

Does Britain belong to Europe?

The past few weeks of dispute about financial policies have confirmed Britain's posture as the odd man out of Europe. Perhaps that is what the British really want.

THE British appear to be digging themselves into a geographically untenable position over their dealings with their European neighbours and among themselves. So much appears to be the only tenable inference from last week's prolongation of the dispute, between the prime minister and her chancellor of the exchequer about the conduct of international monetary policy. Not that last week's instalment of the dispute was any more overt than that preceding the budget in March. The British government naturally denies that there can be a serious dispute at such an august level, and the public signs are all the other way. Last week, for example, Mrs Margaret Thatcher was telling anybody who would stop and listen that Mr Nigel Lawson, the chancellor, is an excellent fellow (or, at least, a "brilliant" one), while Lawson did nothing more provocative than deliver an academic address on the benefits of a free market containing a few sentences implying that governments should sometimes intervene in markets (currency markets in particular).

That was the immediate issue last March, when the chancellor sought to link the British currency more securely with the West German, and when the prime minister believed that the cost of tying the two together would be outrageously high. Like the most interesting quarrels, both may have been correct. In March, when sterling, the British currency, seemed likely to rise too quickly and too far, Mrs Thatcher's ready-reckoning led her to conclude that preventing water running uphill in this way would require a great quantity of sterling to be sold for then unwanted dollars or other denominations, probably without long-term benefit but probably at great loss. The chancellor's view was that high-priced sterling would undermine the recently restored competitiveness of British industry, to the general discomfort, short and long-term. Complicated questions often have this attribute of seeming reasonable from all sides. In the event, the British dispute was caused by a sudden collapse in the strength of sterling (did the quarrel contribute to it?), since when there has been a revival of anxiety about inflation, a countervailing sharp increase of interest rates and now, again, the prospect of more trouble in keeping sterling stable.

Difficulties

Although, on the face of things, these money troubles are arcane, they are proxies for much more deep-seated difficulties in the relationship between Britain and its European neighbours. The immediate connection is the system for regulating European currencies called the European Monetary System, to which Britain has chosen not to belong. It is an open secret that the British chancellor of the exchequer wants Britain to belong as do many other members of the government. The case for membership is not merely the currency stability that follows but the financial disciplines that stability requires. The case against is that in such a system, the cost of stability could be greater than that of fluctuations, while the financial disciplines are but infringements of sovereignty named differently. At least for the time being, there is no prospect of Britain joining the system.

That is the long-term paradox. Like the other members of the European Communities, Britain is committed to being part of a thoroughly common market from 1 January 1992, essentially an

attempt to win the commercial benefits of the United States without a federal constitution. The British seem quite eager to enjoy the fruits of this liberalization of the market, but they do not want the inevitable obligations: a particular issue that has stuck in the British government's throat is that the creation of a truly common market entails uniform rates of value-added tax. So, as on many other occasions in recent decades, Britain is a half-hearted member of a European organization (the European high-energy physics laboratory or CERN, the European Space Agency, the European framework programme of research run from Brussels, are some examples). This is a recipe for the worst of all worlds: membership brings obligations, but mean diffidence means missed opportunities for moulding international institutions in an acceptable way.

For a government as disaffected as the British, it is never difficult to find good logic for hanging back. The European Monetary System will not indefinitely make sense without a European reserve bank (which prospect is now being studied). Can there be such a thing as a European foreign policy without European defence forces? And can there be a European policy on the encouragement of useful science without common mechanisms for supporting basic research? The need cannot indefinitely be overlooked.

Although the presenting symptoms of the present quarrel within the British government are fiscal and monetary, the consequences will thus not leave the professional community indefinitely untouched. One danger for the British and for the rest of the European Communities is that the mainland and offshore parts of Europe will get so far out of step in the coming decade that important opportunities for the more efficient division of labour will be neglected. One illustration of the difficulties that may arise is last week's decision in Britain not to continue spending on fast reactor research (see next page) at a level comparable with that required to sustain effective membership of the European partnership which the government agreed to join only five years ago. Now there seems also to have been trouble about the scale of British participation in the thermonuclear fusion programme. While it is permissible for a member of an international partnership to jib at some proposals for common action, it is difficult to jib at every proposal without, at some stage, letting membership lapse. But is that what the British want?

Maybe it is. In the past decade and more, Britain seems to have been hankering after what can only seem a different path of social development from the mainland partners. Where else in Europe is it considered virtuous to spend time and energy in centralizing control of publicly supported universities, at the same time reducing their capacity, when there is plainly emerging a chronic shortage of new skills of the kinds that universities exist to foster? Where else is the government's budget in surplus while people who wish to go on holiday cannot do so undelayed for lack of adequate air traffic control, and cannot make their way by road instead for lack of public investment? Even the new channel tunnel, due to open in 1992, will be unapproachable by fast train from the British side until many years after it has been finished. It seems odd behaviour. □



Fast reactor move hits European collaborations

- British government eschews the long term
- Atomic Energy Authority looks abroad

London

DRASTIC cuts in development of fast reactor technology in Britain have cast doubt over Britain's role in collaboration with its European partners. The British government last week announced a cut in the present fast reactor research and development from £105 million to £10 million, saying that public funding was not justified because fast reactors would not be commercially viable for 30–40 years. But the government maintains that, with a core programme of research, Britain will continue to contribute to the development of fast reactor technology at an economic cost and this will allow continued collaboration with Britain's European partners.

The UK Atomic Energy Authority (AEA) is not so optimistic. It says that to continue collaboration on research, which began in 1984 with an intergovernmental memorandum of understanding, would be unfair to Britain's partners because the basis for collaboration has been the exchange of intellectual property, and now that Britain's contribution is likely to cease, other countries should not have to subsidize the British programme.

Britain also had an important role to play in a programme of collaboration with the electricity utilities of France, West Germany, Belgium, Luxembourg and the Netherlands to build an experimental fast reactor, probably in France, with a capacity of over 1000 MW, and at a cost of around £2,000 million. A reprocessing plant was to be built to process the fuel, probably at Dounreay in Scotland. Britain's role in the reactor programme was in design. Now the programme may be slowed down as the group decides which country should take on the design of the reactor.

In Britain, the cuts bring a wave of redundancies and cast doubt over the future of the AEA. Over the next two or three years, the authority will shed around 2,500 of its 13,000 staff. And more redundancies may follow. Most of the jobs lost will be at the authority's plants at Harwell (Oxfordshire), Risley (Cheshire) and Windscale (Cumbria). At Dounreay in Scotland, the centre for Britain's fast reactor programme, there will be no job losses in the immediate future. The prototype fast reactor, which accounts for 800 jobs, is due to close in 1993–94 and the associated reprocessing plant three years later.

John Collier, chairman of the authority, feels that closure of the prototype is unjustified. "It comes just at the time when our fast reactor is working better than before", he says. The reactor burns fuel more efficiently than expected, and this year the burn-up target of 20 per cent was achieved.

When the programme began in the 1960s, a future shortage of fossil fuels was expected to push up the price of uranium, which would have favoured fast reactors as, unlike thermal reactors, they run on plutonium. But Britain is still self-sufficient in fossil fuels and, with the discovery of rich seams of uranium, prices have dropped.

The AEA, however, thinks that commercial viability of fast reactors is nearer than the government predicts. And by abandoning a technology in which Britain is a world leader, and has already invested £3,500 million, Britain may miss out on a potentially large export market. Last year the AEA, anticipating future cuts in government funding, began a search for replacement contracts. And last week an agreement was signed with the Sumitomo corporation of Tokyo, which will act as agent for the authority in Japan. The authority predicts contracts worth millions of pounds from the agreement. It will now intensify its attempts to find alternative funding and will appoint 20 marketing specialists to speed up the process.

Christine McGourty

■ **T**HE future of British research into nuclear fusion, over which Britain has been haggling with ministers of the European Community, is in doubt as a result of the government cuts in fast reactor research, which reflect a policy of withdrawal of public funding for projects with only long-term commercial benefits.

An agreement has finally been reached on a budget for the European fusion programme after Britain had angered its fellow members by insisting that more than 10 per cent of the budget should be held over until 1992. But a compromise was reached; 50 million ECU (European Currency Unit; 1 ECU = £0.69) of the budget of 735 million ECU will be held over. The agreement will be formally approved at the next meeting of the European Parliament.

Funding for fusion research in Britain is secure only until 1992, when the Joint European Torus at the Culham Laboratory in Oxfordshire will be completed. **C.McG.**

Fast breeder generates controversy

Munich

THE fast-breeder reactor at Kalkar has generated nothing but controversy after 16 years of construction, 18 planning permits and DM7,000 million of investment. In mid-July, Environment Minister Klaus Töpfer had to rush back from vacation to settle a cabinet crisis over the reactor. Töpfer (Christian Democrat, CDU) is now threatening to bring the Social Democratic (SPD) government of the Land of North Rhine-Westphalia (NRW) before the federal constitutional court in order to force it to resume the reactor licensing procedure.

The 300-MW plant is located 70 km west of Düsseldorf, not far from the Dutch border. Dutch and Belgian electricity companies are partners in the project and would have to be repaid about DM1,300 million if the project were abandoned.

Kalkar was begun in the 1970s by an optimistic West German government which saw it as a step towards completing the nuclear fuel cycle. The Federal Minis-



try for Research and Technology (BMFT) paid the major share of the construction costs and is now losing DM10 million per month because of the delay.

North Rhine-Westphalia's government's enthusiasm for the project vanished after the Chernobyl accident in 1986. Critics claim that Kalkar will be out of date by the time it goes into operation in the early 1990s; that it will not 'breed' new plutonium anyway and is therefore unnecessary; and that there is already ample generating capacity.

BMFT spokesman Holm Kilbert asserted that Kalkar can still play a valuable role in research on nuclear power. Japan began building a fast breeder quite similar to Kalkar just a year ago, he said, so Kalkar "can't be so wrong". West Germany is part of a European consortium to improve nuclear power plants. West Germany would disappoint Great Britain and France if it cannot put Kalkar into operation. **Steven Dickman**

Biological weapons research opposed

Washington

A NEW campaign against the military use of biomedical research was launched in the United States last week with the publication of a pledge signed by 500 prominent researchers.

The pledge, which commits signatories "not to engage knowingly in research and teaching that will further the development of biological or chemical weapons", is the work of the Committee for Responsible Genetics (CRG)*, a Boston-based pressure group.

CRG hopes that the existence of the pledge will help stiffen scientists' resolve not to take money for US Department of Defense biological research projects. More money has been on offer since the Reagan administration quadrupled the budget of the Biological Defense Program between 1983 and 1986.

Both the United States and the Soviet Union are signatories of the Biological Weapons Convention of 1972, which bans the research development, stockpiling and use of biological weapons. But the Reagan Administration has revived a programme which had been cut by President Richard Nixon in 1969.

According to Richard Novick, director of the Public Health Research Institute in New York and a member of CRG's advisory board, the reasons Nixon decided to abandon biological weapons research remain sound. Biological weapons have no strategic or tactical value in a conflict between the major powers because their spread cannot be effectively controlled. Any use by one side would be quickly detectable and, if threatening, could provoke a nuclear response from the other side.

Biological weapons make military sense only in covert operations against third-world countries, he says. Water supplies could be contaminated with pathogenic micro-organisms and crops destroyed by release of pests. Third-world countries do not have the resources to prove that biological warfare is being used against them.

Any defensive research by the United States always runs the risk of being seen as preparation for offensive operations and could trigger an arms race, according to Jonathan King, professor of molecular biology at MIT and a CRG board member. And there is the additional risk of an accident. At the Dugway Proving Ground, Utah, which the Reagan administration is planning to reopen as an aerosol test facility (see *Nature* 333, 106, 1988), thousands of sheep died in the 1970s when the wind direction changed during a nerve gas test.

Alan Anderson

*Copies of the pledge are available for signing from CRG, 186A South St, Boston, MA 02111-2701.

ABRC calls for more money to boost science

London

THE British government's advisers on the science budget are asking for more money to continue restructuring academic science and to pursue new scientific opportunities. In a report to the Secretary of State for Education and Science, issued last week, the Advisory Board for the Research Councils (ABRC) calls for an extra £97 million in 1989-90 and an extra £131 million and £151 million in the subsequent years. Present spending plans call for £730 million for 1989-90, rising to £750 million in 1991-92, implying a reduction of at least 3 per cent over the next three years after allowing for inflation.

Last year, ABRC asked for £130 million for 1989-90 and £160 million for the following year, but won increases of only £65 million and £48 million for those years. The board now says that was insufficient, either to allow for the reshaping of the science base or to avert a reduction in the amount of research.

Recent data on scientific publications and citation indicate that Britain's share of world scientific output and its influence are declining, says the report, at a time when industry is becoming increasingly science-dependent. Sir David Phillips, chairman of the board, says that although industry should pay for applied research and development, public support is vital to its underpinning of basic science.

Of the total estimated cost of restructuring of £179 million over the next three years, £105 million will be for setting up 30 interdisciplinary research centres. The remainder would aid faster restructuring of the research councils to make them more efficient and effective.

An extra £200 million is needed over the same period, says ABRC, to grasp a range of new scientific opportunities in areas including computer science, quantum optics, the human genome and atmosphere and marine pollution. The remainder would go towards relieving manpower problems, buying scientific equipment and giving increased support to high-quality grant applications. Phillips said the shortage of manpower in research is "absolutely frightening".

Fears over manpower problems were reinforced last week with the publication of a survey of the destination of graduates, by the Association of Graduate Careers Advisory Services. It showed that, though the market for graduate employment is booming, there was a fall last year in the proportion of graduates going into science and engineering research and development. From a peak of 17.4 per cent in 1984, the figure fell to only 14.5 per cent last year.

ABRC stresses that money earmarked by the government for special cases should not divert the finances of the research councils. The board says it will need an extra £42 million over the next three years to meet the costs of four programmes the government is at present discussing: contributions to the European particle physics laboratory, CERN; the British Geological Survey; the Medical Research Council's AIDS programme and the British Antarctic Survey. Christine McGourty

Science advice

London

BRITAIN'S Parliamentary and Scientific Committee is setting up a new body to provide members of parliament with independent scientific advice on policy issues. The advisory board to the new body, called the Parliamentary Science Foundation (PSF), met for the first time last week



PSF chairman Sir Ian Lloyd.

under its chairman Sir Ian Lloyd to discuss the selection of a director, who it is hoped will be in post by December.

The board says an independent advisory board is vital if MPs are to make informed judgements in areas of public policy influenced by science. Lloyd says that at present MPs risk being the victims of lobbying by interested parties, and are unable to arbitrate between differing scientific judgements.

The new body is to be modelled primarily on the United States Congressional Office of Technology Assessment, which has a full-time staff of 140 and a budget of \$17 million. But Britain's version will be somewhat smaller, costing about £2 million a year, says Lloyd. So far, the board has raised £177,000 from companies, scientific institutions, universities and individual donations.

When PSF is granted charitable status, a major national appeal will be launched, and parliament will be asked to match the support pledged from donations. The board hopes the new body will be in operation by the end of 1989. Christine McGourty

Leak jeopardizes shuttle launch

Washington

A LEAK in an oxidizer fuel-tank on the space shuttle *Discovery* threatens to delay the planned September resumption of shuttle flights by as much as two months.

Engineers and technicians at the National Aeronautics and Space Administration (NASA) Kennedy Space Center have postponed until after a test firing of the shuttle's main engines on 28 July the decision whether to attempt to repair the leak on the launch pad or to return *Discovery* to the Orbiter Processing Facility.

NASA engineers have localized the leak to a seal on a vent line above the oxidizer tank for the reaction control system, a set of thrusters used to control the spacecraft's position in orbit. But the seal cannot easily be reached, in its location inside the left-hand pod housing the manoeuvring system at the base of the orbiter, so that it is unclear whether it can be repaired on the launch pad. NASA transferred *Discovery* to the launch pad on 4 July.

Finding the leak forced a two-day delay in preparations for this week's firing of the main shuttle engine, intended to prove its readiness for flight. That 20-second test is one of last crucial steps needed to declare the space shuttle orbit-worthy. Another hurdle that must be cleared is a final test firing of the newly designed solid rocket booster now scheduled for next month.

If *Discovery* must be returned to its hangar, there will be at least a two-month delay before the flight can be rescheduled.

Joseph Palca

Progress at CERN

London

FINANCIAL problems aside, things are going well at CERN, the European laboratory for particle physics. Two weeks ago, a beam of four positron bunches was successfully injected into the new Large Electron-Positron Collider (LEP) and transferred into the first 2.5 km of the tunnel. In what was the last major milestone before LEP is switched on next year, the whole injection system was found to work perfectly.

When completed, LEP will be the world's most powerful electron-positron colliding beam machine and will give particle physicists worldwide the facilities for performing experiments at energies of 200 GeV. LEP consists of a water-cooled, lead-shielded vacuum tube surrounded by magnets. Electrons travelling clockwise and positrons circulating anticlockwise will be kept on track by the magnets inside this same vacuum chamber. Construction began on LEP in 1983 and the total cost is around £400 million. Christine McGourty

Caltech seeks to improve its seismic monitoring

Berkeley

WHEN the magnitude 5.9 Whittier Narrows earthquake jolted Los Angeles last October, seismometers near the site went off the scale, requiring California Institute of Technology seismologists to rely on readings from stations hundreds of miles away. Within a few years, Caltech seismologists hope to be able to monitor earthquakes as large as magnitude 7 from nearby stations, thanks to a new southern California seismic network, called Terrascope, for "terrestrial telescope".

Over the past five years, Caltech seismologists have realized that their current seismic array of more than 250 conventional analog or digital seismometers is "woefully obsolete", says Don L. Anderson, director of Caltech's seismological laboratory. Data from the seismometers, some of which have been

...and when the seismometers fall off the wall, it's a magnitude 6...



in operation for 60 years, can take months or years to analyse. The process is slowed by tedious analog-to-digital data conversion, and complicated by the limited dynamic range of conventional seismometers, which allow magnitude 1 or 2 earthquakes to slip by undetected, while those of magnitude 5 or greater send the instruments off scale.

The Terrascope will allow seismologists for the first time to make continuous observations of the complete process of earthquakes, including minor foreshocks and aftershocks, and the slow creep that occurs between earthquakes. When completed, the Terrascope will consist of a \$4.2 million array of at least 10 seismometers with a dynamic range 10,000 times greater than that of conventional instruments, allowing detection of a broad range of earthquake intensities. They can also detect a much greater range of frequencies of vibration than conventional seismometers.

With the help of continuous digital recording and fast computers, the array will provide data in real time, even during the earthquake. A computer database will allow comparisons with earthquakes in the area over the past 60 years.

The Terrascope stations will also be linked to the Global Positioning Satellite network, which will provide position data precise enough to allow tracking of the slow (several centimetres per year) movement of the tectonic plates that occurs between earthquakes.

A prototype station, a collaborative effort between Caltech, the University of Southern California, Harvard University and the US Geological Survey, has been operating for six months in the Pasadena hills. It has remained on scale through nearby magnitude-5 earthquakes, said Anderson, and based on its performance Caltech is ready to build the array.

The Terrascope's seismic stations will use 'off-the-shelf' equipment, said Anderson, including a Swiss sensor and a data logger developed by a Harvard graduate student. Anderson said the whole array could be completed within three years, once funding is secured. Marcia Barinaga

Thin films carry more current

Tokyo

IN an important step on the way to practical applications of high-temperature superconductors, researchers at Sumitomo Electric Industries in Japan have made a polycrystalline thin film of bismuth-strontium-calcium copper oxide with unusually high current-carrying capacity.

The bismuth-based superconductor, developed in Japan earlier this year, has been the object of intensive study, whereas in the United States the equally new thallium-based materials are the main attraction. But until now the current-carrying capacity of bismuth superconductor has reached no more than about 100,000 amps per cm² in thin films, too low for electronic applications.

Sumitomo researchers say they have increased the current-carrying capacity by an order of magnitude to 1.9 million amps per cm² at liquid-nitrogen temperature by producing a polycrystalline film in which the crystals are aligned so that the copper-oxide planes, along which current flows most easily, are parallel to the film.

Comparable critical currents have previously been reported only for single crystals of rare earth superconductors (for example, see *Nature* 328, 98; 1987). Polycrystalline films, potentially more useful in electronics because they are not limited by crystal size, usually show much lower critical currents because of the effects of crystal boundaries. David Swinbanks

Human Frontiers Program to go international in Switzerland

Tokyo

AFTER years of deliberation, Japan's Human Frontier Science Program will finally begin today (28 July) with the launch of a pilot grant programme. More important, the Science and Technology Agency (STA) and the Ministry of International Trade and Industry (MITI) are shortly to propose that the full-scale programme be launched next year, backed by an international foundation to be established in Switzerland.

The move to establish a foreign scientific foundation designed to support international basic research is without precedent in Japan. Cooperation will be sought from scientists of the 'summit' nations (Japan, United States, United Kingdom, Canada, France, West Germany, Italy and the European Community).

Tateo Arimoto, director of planning in the STA's Science and Technology Policy Bureau, says the foundation and programme will be modelled "flexibly" along the lines of the European Molecular Biology Organisation. The foundation will be

entirely funded by Japan in the initial phase; STA and MITI will request "a minimum of \$20 million" to start things going in fiscal year 1989. The first international grants should be available from the Swiss foundation in about October 1989.

Although other summit nations will not contribute funds in the initial stage, Arimoto says that the US National Science Foundation and the European Community will provide experienced personnel for the administrative secretariat of the programme. And scientists from all summit nations will participate in the governing council and peer review committees of Frontiers. Ultimately, Arimoto hopes to reach a multilateral agreement for the establishment of an organization similar to the World Health Organization and funded by all summit nations.

A committee of 20–25 summit scientists will convene the first of several meetings this autumn to discuss the way Frontiers will begin. And four or five international workshops on the brain, the human genome project, molecular recognition and non-invasive techniques for scanning biological functions will be held in the period up until the end of fiscal year 1988 (March 1989); the agency has set aside ¥300 million (\$2.25 million) for them in this fiscal year.

Meanwhile, MITI today announced the availability of international grants under an 'experimental' Frontiers programme (see the 'Classified' section, page 22). Five

grants worth up to ¥30 million (\$225,000) will be awarded in this fiscal year to "international joint research teams" composed of researchers from institutions in at least three summit nations, one of which must be Japan (this latter condition may be dropped in the full-scale programme).

Grants for research into 'high-order brain functions' or 'molecular recognition and response', will be for up to two and a half years and will be administered by the New Development Organization (NEDO).

NEDO is a a semigovernmental organization that was set up jointly by MITI and private industry in 1980 to build pilot plants for tapping alternative energy sources. But under legislation passed last spring, the organization will be re-organized to incorporate many of MITI's major research and development programmes, including Frontiers (see *Nature* 333, 4; 1988). Other Japanese ministries may join Frontiers from fiscal year 1990.

The Ministry of Education, Science and Culture (Monbusho) was apparently shocked when MITI was awarded ¥170 million (\$1.3 million) for distribution of the international grants this year. Although the amount of money involved is relatively small, basic research and international collaboration in science has traditionally been the preserve of the education ministry.

According to a Monbusho adviser, the ministry has "reluctantly" entered into negotiations with the STA and MITI and is still maintaining a 'wait-and-see' attitude. But STA and MITI officials are now confident that they have the full backing of Prime Minister Noboru Takeshita and will win the funding they want.

David Swinbanks



Yusacheslav Lebedev acknowledges his welcome to the US Air Force base at Greenham Common in England. Lebedev was leader of the 20-member Soviet inspection team that last week carried out the first inspections under the US-Soviet treaty scrapping intermediate-range nuclear weapons in Europe. The Tomahawks at Greenham Common and other European bases are to go, but the exclusion of sea-launched cruise missiles from the INF agreement raises doubts about the future of the Strategic Arms Reduction Treaty. See this issue, page 307.

Creeping towards reform in Australian higher education

Sydney

AUSTRALIA'S higher education system is being reformed a little at a time. In a 'pay for productivity' deal, academic researchers have agreed to allow the number of untenured senior academic posts to increase from 3 per cent to around 10 per cent of total posts.

Also part of the package is the adoption of uniform provisions for the dismissal of academics on grounds of unsatisfactory performance, and an early retirement scheme. Until now, serious misconduct has been the only basis for firing an academic. The early retirement scheme should see 1,000 academics retire over the next three years.

These moves are all among reforms designed to increase the flexibility of higher education institutions that was advocated in a green paper by the Minister for Education, Employment and Train-

ing, Mr John Dawkins (see *Nature* 330, 592; 1987).

Academics will receive a 4 per cent pay rise that will cost the government an extra A\$46 million in the next year. No extra money will be forthcoming in succeeding years, since by then the 'productivity gain' is expected to have worked its way through the system. Further changes to academic's rights may come from a log of claims submitted to arbitrators by university vice chancellors. One contentious request is that some academics may be released on grounds of redundancy.

Although academic unions maintain their commitment to tenure, Professor Ralph Hall, vice president of the Federation of Australian University Staff Associations, says that if the vice chancellors can win the power to make academics redundant, "tenure will become virtually meaningless."

Charles Morgan

GAO report vindicates Teller but critics disagree

Berkeley

PHYSICIST Edward Teller was not alone among weapons scientists in believing that the X-ray laser could become the backbone of the Strategic Defense Initiative (SDI), nor is there a scientific consensus that he misled government officials, according to a long-awaited report by the US General Accounting Office (GAO)*.

But critics of the report say its findings are inaccurate and blame politics for clouding its conclusions.

The report was requested by US Representative George Brown after Roy Woodruff, a physicist at Lawrence Livermore National Laboratory (LLNL), accused Teller of misleading government policy-makers. Woodruff also claimed that the former LLNL director, Roger Batzel, would not allow him to present his criticisms in writing. Woodruff resigned as associate director for defence systems in October 1985, saying he would not be responsible for fulfilling Teller's promises. Later he filed a personnel grievance with the University of California (UC), which oversees the laboratory, for what he felt were reprisals, based on his opposition to Teller. He won his grievance, and was reinstated as associate director for weapons verification research in December 1987 (see *Nature* 329, 751; 1987 and 330, 594; 1987).

Among its findings, the GAO report states that the "official channel" at LLNL had made statements about the X-ray laser similar to those made by Teller and

challenged by Woodruff. And, while Woodruff was asked not to send his written objections to Teller's statements to government officials, the report notes that he did present those views verbally to a number of top government officials, including presidential science adviser George Keyworth and chief arms control negotiator Paul Nitze. The report characterizes Keyworth as a knowledgeable physicist who knows Teller as a "technical optimist", and did not need clarification of Teller's views. It also suggests that Teller was not alone in his views, but that scientific opinion about the feasibility of the X-ray laser covered the whole spectrum.

Current LLNL director John Nuckolls, not available for comment on the report, said through a laboratory spokeswoman that it was "the best that could be expected, given the complexity of the situation". Woodruff also gives the report a lukewarm review, calling it "marginally acceptable". By documenting statements and letters by Batzel, Teller and Teller's protégé Lowell Wood, he says the report confirms "that everything I ever said is absolutely correct". Woodruff dismisses the report's conclusions as not supported by the facts, and contends that the report was "deliberately made confusing, because it is a hot political issue". Nevertheless, he says the report now allows him to discuss the real reason he resigned, which was the previously classified plan for a multiple-beam X-ray laser, known as Super Excalibur, characterized by Woodruff as a last-ditch effort to design a system that would make SDI possible. The report cites a letter from Teller to Nitze that describes Super Excalibur as "a single X-ray laser module the size of an executive desk which . . . could potentially shoot down the entire Soviet land-based missile force, if it were to be launched into the module's field of view". But that, says Woodruff, is like promoting the atom bomb before scientists had any idea of how to split the atom.

Woodruff says the GAO report supports his complaint that Super Excalibur was no more than "pie in the sky", a weapon on paper only, that Teller was promoting with no supporting data, in the hope of influencing the Geneva arms reduction talks. The report quotes a December 1984 letter from Teller to national security adviser Robert McFarlane, in which he said that Super Excalibur "might be accomplished in principle within a few years", as well as a conflicting statement by Batzel to the House Armed Services Committee in February 1986 that "there are no data at this stage of the game" that would support

the feasibility of Super Excalibur.

The report also addresses the controversy over Excalibur, a more modest proposal for a single-beam X-ray laser which would not be able to shoot down an incoming missile barrage, but rather whose major application would be as an anti-SDI weapon. The report notes Woodruff's objections to Teller's assertion that Excalibur was ready to enter engineering phase. The brightness of the laser had not been confirmed, according to Woodruff, and critical scaling experiments remained to be done. The report notes that George Miller, who succeeded Woodruff as associate director for defence systems, agrees with Woodruff that Excalibur was not ready for engineering then, nor is it today.

Woodruff, while satisfied that the facts have been aired on his disagreements with Teller, says the real issue is the role of the University of California as overseer of the laboratory in such a dispute. Woodruff says he is still living under the stigma of being the "whistle-blower", and fears he may be forced out of his new job. He complains that UC has not defended his academic freedom to balance Teller's scientific claims, and that UC president David Gardner has never spoken to him nor agreed to investigate his complaints.

The university, for its part, characterizes Woodruff's employment complaints as a personnel issue, to be handled internally by the laboratory. Gardner has pointed out that Teller, as a retired LLNL director, was not an official spokesman and was free to express his personal views in writing, while Woodruff, as an associate director, would, if he commented in writing, be seen as giving the official stamp of approval to a scientific opinion.

The California legislature included in this year's UC budget a request that the university increase its oversight of the Livermore and Los Alamos weapons laboratories, with one specific goal being to ensure that research is technically sound and not misrepresented to government officials, and that dissenting views be presented.

At their 15 July meeting, the UC regents chose not to adopt the budget request, but rather to initiate a study to evaluate the need for greater oversight. According to a UC spokesman, an increase in university oversight is unlikely, as Gardner believes that it would conflict with the jobs of the laboratory directors.

In a statement to the regents, Gardner reiterated his philosophy that management of the laboratories should parallel the management of individual UC campuses, with relative autonomy given to the directors, as it is to the chancellor of each campus.

Marcia Barinaga

UK record network

London

BRITAIN urgently needs a national network of biological records to bring cohesion to the present inefficient and ramshackle system, according to a committee of the Linnean Society, which issued its report two weeks ago. The chairman of the committee, Professor Sam Berry, president of the British Ecological Society, urged the need of a national system so that planners can make informed decisions on land use, especially with the prospect of changes in the countryside as agricultural land is taken out of production and as cities exert pressure on surrounding green belt areas.

The committee makes four main recommendations: a national coordinating commission; a national network of biological records centres linking and supporting local centres; a national collative and interpretative unit; and a central data store.

The total cost of the proposals would be between £300,000 and £1 million.

Christine McGourty

*Strategic Defense Initiative Program: Accuracy of Statements Concerning DOE's X-Ray Laser Research Program. US General Accounting Office Report, June 1988.

Intensified search for energy alternatives in West Germany

Munich

RESEARCH Minister Heinz Riesenhuber (Christian Democrat) announced in Bonn last week that West Germany's investment in renewable energy research — DM260 million for 1988 — was the largest in the world "in both absolute and relative" terms.

But many of those who support increased use of renewable energy resources are still not happy with West Germany's centre-right government. That is because the government is holding firm to nuclear power as an 'interim' energy source along with the commitment to renewable energy research. Thirty-eight per cent of West Germany's total energy use is supplied by nuclear power despite growing resistance from the press, the public and several *Länder* governed by the opposition Social Democrats (SPD). Renewable energy sources such as hydroelectric, solar and wind power account for 2.4 per cent of West Germany's energy consumption.

Critics also say that the government has failed to use subsidies effectively to hasten the introduction of new energy sources. One government official said that BMFT wanted more money for such subsidies but the request was refused by Economics Minister Martin Bangemann (Free Democrat, FDP). The FDP, the smallest partner in the governing coalition, opposes government subsidies.

The cost of energy in West Germany supports the federal government's continued enthusiasm for nuclear energy. According to government figures, one kilowatt-hour from a nuclear power station costs DM0.14, compared with DM0.27 for wind energy, DM3.57 for solar energy and DM0.20 for energy from coal-fired power stations.

A number of SPD-governed *Länder* are aggressively pursuing non-nuclear options and talking about following the decision recently made in Sweden to end reliance on nuclear power in the early twenty-first century. In the northernmost West German *Land* of Schleswig-Holstein, the SPD was given an absolute majority in elections on 8 May; its platform promised a shutoff of nuclear power.

Hamburg's plans to eliminate dependence on nuclear power by 1996 received a setback last year when the SPD was forced to form a governing coalition with the FDP. Hamburg has nevertheless commissioned an independent report on the economic consequences of the elimination of nuclear power which is due out in September. Hamburg relies on nuclear power for 90 per cent of its electricity.

In Saarland, a coal-mining region which uses no nuclear energy, and Schleswig-

Holstein, the SPD is working on the coupling of coal-fired electricity plants to heat exchangers which then carry the excess heat to customers through long-distance heating pipes.

Such a procedure could increase to 90 per cent the recovery of energy from coal. Traditional coal-fired power plants use only 35 per cent of the energy and release the excess heat to the atmosphere. Under such a plan, numerous small power plants

would have to be built near populated areas. A pilot plant is already operating "very economically" at Flensburg, said a spokesman for the Schleswig-Holstein energy ministry.

Both *Länder* intend to encourage energy conservation among the population. Remarkably, the first pilot study on the effectiveness of electricity conservation in West Germany is being carried out only this year in Saarland. The government has set up an 'energy agency' to finance energy-saving home improvements and billing procedures are being changed to encourage savings.

Steven Dickman

Congress provides a rundown on US biotechnology

Washington

LAST week's release of a report by the US Office of Technology Assessment (OTA) was the highlight of a week of congressional attention to biotechnology, which featured a day-long conference and a hearing on a bill to set up boards to enhance the competitiveness of the industry.

The report, entitled *US Investment in Biotechnology*, is the second to examine US efforts both to encourage the growth of the biotechnology industry and to remove any federal obstacles in its path since the biotechnology industry began in earnest in the early 1980s.

In the new report, the OTA finds that the US biotechnology enterprise is roughly equally underwritten by the federal government, which spends \$2,700 million annually on basic and "generic applied" research, and private industry, which invests nearly \$2,000 million.

For the largest sector of the biotechnology industry — that developing human pharmaceuticals — the investment is judged sufficient, with the exception of research in protein chemistry and drug delivery systems. But the key areas of agricultural biotechnology and biological waste disposal techniques are suffering from a lack of money and from regulatory barriers governing environmental releases.

State governments have become increasingly involved in attempts to foster biotechnology companies. Although their investment is only one-sixteenth of the federal government's, 33 states have either formally organized biotechnology centres to draw in large companies and incubate start-ups, or provided tax and training incentives.

Not surprisingly, biotechnology companies thrive best in the states with the strongest university research programmes in the biological sciences. OTA reports that the industrial contribution to aca-

demic research is four to five times greater in biotechnology than in other fields.

The close relationship of universities and biotechnology companies has so far not led to a corruption of basic research interests or diverted academic scientists from their student education responsibilities, but OTA says collaborations may need monitoring.

The OTA report suggests that additional financing may be needed for new biotechnology companies, perhaps in the form of additional small business grants. OTA also found that the tax reform law passed in 1986, which reduced the tax credits for research and development and which abolished the lower tax rates for capital gains from the sale of research and development limited partnerships, were harming industry.

The OTA report did not bear out some of the dire predictions made by the agency in a similar report issued in 1984. The biotechnology industry has not experienced a 'shakeout' eliminating all but the largest and most profitable companies, and there has been no serious lack of funds for applied research.

No critical shortage of personnel trained in the biological sciences and in bio-process engineering has occurred either, with the exception of microbial ecologists, who are in tight demand.

The congressional hearing held last week discussed bills that would establish national policy commissions and ethics boards to steer US biotechnology activities, including the project to sequence the human genome. A bill has already been passed in the Senate, and legislation in the House of Representatives is pending.

Biotechnology trade associations welcomed the creation of boards intended to foster their industry, but administration officials testified that another layer of bureaucracy was unnecessary and that the legislation as drafted may in fact be unconstitutional.

Carol Ezzell

Defeat on education bill

London

BRITISH academics, flushed by several victories in the long battle with the government over the education reform bill, suffered a defeat last week. A House of Lords' amendment on university funding was reversed in the House of Commons. As the bill now stands, grants made to universities will be subject to "terms and conditions" laid down by the new Universities Funding Council (UFC).

In the bill's original form, the word 'payments', not 'grants', was used. The wording was changed in the House of Lords to "grants, specifying particular obligations and subject to general guidance" (see *Nature* 334, 7; 1988). Academics saw this change as necessary if universities were to retain some independence.

The government has now agreed to retain the word 'grants', but has reinserted the 'terms and conditions' clause. The Secretary of State for Education and Science, Mr Kenneth Baker, says this is essential to allow the ear-marking of funds for specific purposes, such as the current restructuring of university departments and in order to ensure proper accountability. But the government denies that the UFC will have more wide-ranging power than the University Grants Committee (UGC) now has.

At present, UGC advises the government on the annual grant to each university. UGC can offer 'advice' on how money is spent, but that can be ignored on the understanding that future grants could be cut. Yet this bill empowers UFC to demand repayments of grants with interest if the conditions of the payment are not met.

In view of other concessions by the government, the Committee of Vice-Chancellors and Principals (CVCP), has agreed to accept the clause as it now stands, but the Lords may yet oppose the clause this week, thus delaying the bill. But a defeat in the House of Lords seems unlikely. CVCP remains concerned about the interpretation of the clause and the proportion of university finances to be ear-marked. These issues will be discussed with the government.

The concessions by the government on higher education include the acceptance of a statutory guarantee of academic freedom in the face of the abolition of tenure; academics will have "freedom within the law to question and test received wisdom and to put forward new ideas and controversial and unpopular opinions without placing themselves in jeopardy of losing their jobs or privileges".

Christine McGourty

OTA produces a consumer's guide to new rockets

Washington

It is a rather specialized market, but the Congressional Office of Technology Assessment (OTA) has produced a buyer's guide for those considering the purchase of new launch systems to maintain and enhance the US presence in space.

The OTA report argues that before deciding on a launch system, Congress must decide where the US space programme is headed. With no expansion of space activities, the current launch fleet of the space shuttle and expendable launch vehicles (ELVs) — including the Titan II, III and IV, Scout, Delta and Atlas/Centaur rockets — can reasonably be expected to place up to 956,000 pounds of payload into low Earth orbit. But for an expanded space effort, including deployment of a space-based ballistic missile defence, full deployment of the space station and human exploration of Mars, the current capabilities are insufficient.

One option is to enhance the existing launch fleet. The National Aeronautics and Space Administration (NASA) is funding work on an advanced solid rocket motor that would replace the solid rocket boosters currently used on the space shuttle. Liquid rocket boosters are also being developed that would be more powerful than the solid fuel rocket now used. Both systems would add approximately 12,000 pounds the shuttle's lift capabilities.

By making the shuttle's external fuel tank out of an aluminium-lithium alloy, another 12,000 pounds could be added to the maximum cargo load. OTA estimates that these improvements, coupled with enhanced performance from existing ELVs and slightly higher launch rates, could raise annual launch capacity for low-Earth orbit to 1,600,000 pounds. OTA considers this option a 'best buy' if the US space programme is to stay at current or slightly enhanced levels of activity.

But other options are definitely on the drawing board. NASA, along with some allies in Congress, sees development of Shuttle-C as a smart way to go. Shuttle-C is an unpowered version of the current shuttle. It would use the same solid rocket boosters and external fuel tanks, but in place of the orbiter there would be a payload shroud with either two or three main engines. Shuttle-C would be capable of lifting a 100,000-pound payload into low-Earth orbit. OTA points out that there have been objections to this proposal because it "places all our eggs in one basket", relying on the same technology for piloted and unpiloted flight. OTA says Shuttle-C's strongest selling point is that it

would considerably shorten the time needed to deploy the space station. But it is still not clear that Congress will find the funds to permit the space station to go forward.

Another near-term option is a transition launch system based on current technology, but capable of lifting up to 150,000 pounds into orbit. Congress ordered the Air Force not to pursue this option, in large part because it wanted to forestall its use for an early deployment of a space-based missile defence foreseen by the Strategic Defense Initiative. But OTA says the idea should not be abandoned if Congress wants an increased launch capability by the mid- to late-1990s.

Along with NASA, the Air Force is currently pursuing a project known as the Advanced Launch System (ALS), designed to bring launch costs down to \$300 per pound for low Earth orbit. Richard DalBello, project director for the OTA report, says ALS got off to a slow start because NASA and the Air Force wrangled over how the project should be managed. Now the Air Force seems to be calling the shots, and expects to have a preliminary design review completed by June 1990.

The National Aerospace Plane proposed by the Reagan administration represents a dramatic departure from current launch technologies, requiring 'scramjet' engines that could operate in the atmosphere and space. The OTA report says there are several major technical hurdles to overcome in this project, but the payoff could be a fleet of vehicles capable of taking off from conventional runways and flying directly to Earth orbit.

Although ALS and the aerospace plane are intended to use new technology, there are even more unconventional options being considered for getting payloads off the Earth. A ram cannon uses a barrel filled with fuel that ignites as a projectile passes through it. According to OTA, tests of the ram cannon have propelled a 0.1 pound projectile at 20,000 g to a velocity of 1.25 miles per second, 20 per cent of that needed to reach Earth orbit.

More exotic still is a proposed rocket using anti-hydrogen fuel. Only 35 milligrams of anti-hydrogen would be needed to boost 30 tons into low Earth orbit. One drawback to this scheme is that at current rates of production it would take 19 million years to produce the necessary amount of anti-hydrogen. But with a 10 gigawatt solar-powered orbital production facility, some estimates place production time at only five weeks. OTA is not necessarily endorsing this option at present.

Joseph Palca

Explanation of Benveniste

The following letters are triggered by an article by Davenas et al. (*Nature* **333**, 816–818; 1988).

SIR—The article stating that high dilutions of a protein (containing no protein molecules) are biologically active has all the traditional properties of homeopathic claims: insufficient description of the methods used (what is the source of the 0.1 per cent human serum proteins in the diluent?), suggestive hearsay ("similar results were obtained in other laboratories (Toronto, preliminary results)") and wild statements with no data shown ("our results can be summarized as"; "we can confirm that").

This will turn out to be yet another case of artefacts (or worse) but the harm has already been done; worldwide recognition of this important paradigm of homeopathy by a major scientific journal. The homeopathic industry (supplying for example 15 per cent of all medicines in France) will not care about a short retraction in *Nature* (which will get much less attention in the popular press than the original claims).

Nature should have insisted on an independent confirmation of these results before and not after publication.

RONALD H.A. PLASTERK

Netherlands Cancer Institute,
Plesmanlaan 121,
1066CX Amsterdam,
The Netherlands

SIR—Scientific belief belongs on a flat earth. There is no danger, no threat to science, in the restatement of the drug-diluent paradox — we need only apply the scientific method and then seek the verdict of experience.

In this way the inquiry by me and my colleagues reluctantly judged the placebo hypothesis redundant and homeopathic dilutions active¹. We agree with Benveniste's suggestion that his results relate to ours and we consider that both teams' research represents a development from homeopathic science. Those who adopt the position that there is no such link are standing on ice so thin that even the weight of the author list will crack it — there are at least three homeopathic doctors on it. It may be deep water but we must face the facts that homeopathic physicians introduced and developed this area and that those wishing to examine this 'new' discovery will be well served by homeopathic history. Fifty years before *Nature* had the courage to publish, Boyd began to demonstrate in a series of classical scientific experiments similar on/off effects up to 10⁻³⁶ in biological assays such as mercuric chloride affecting starch diastase². His methods were scientific but his results 'unbelievable' and so they

joined a body of experience accumulated and ignored over the very same 200-year period that you invoke in your leading article (*Nature* **333**, 787; 1988). If we now rise to this challenge, we have nothing to lose. If we prove the observations wrong, we will have exposed homeopathy as one of medical science's greatest misadventures — a folly so massive it will merit study in itself — why have 1 in 4 French doctors prescribed it³? If confirmed, we open up a vista of fundamental discovery and development. If this strikes at the heart of some current scientific models, that will be good for science. Right or wrong, such cross-fertilization has yielded benefit in the past. Some claim it laid the foundation of modern immunology⁴, certainly tangible benefits to medicine accrued when homeopaths proved pollen the cause of hay fever⁵ and introduced low-dose allergen desensitization⁶.

Moving from experiment to speculation, how are we to conceptualize memory in water? (If it is there, then the later stages of reaction with biological systems are easier to imagine with the models of pheromones, receptor sensitivity and biological amplification.) As a point of departure for criticism, I have visualized this first link in the chain as 'liquid snowflakes' modelled from the parent drug — one concentration but with unique biophysical information or pattern, 'seeded' in the new diluent in the reorganizing post-vibrational phase like liquid crystals growing in the order after chaos. But my snowflake analogy melts if we become fixed on the idea of 'shape' — best to say for now that the quality and nature of this ghost in the machine is unknown. And what of sinusoidal change in activity? Might the first generation antagonist grow in dominance through successive dilutions until at a critical threshold (concentration?) it itself becomes the template for a second generation agonist, the relative activity of these opposing patterns alternating in a sinusoidal curve as one continues the process of dilution/vibration — an agonist/antagonist template shift. In this respect, other work by Poitevin, Davenas and Benveniste⁷ shows that certain dilutional points are not just ineffective as agonists but show an opposite inhibitory effect on the action of a second agonist. Similarly, we have shown clinically that although most patients reacting to a homeopathic allergen dilution improve, unpredictably some aggravate and then improve while others only aggravate without subsequent improvement. In clinical practice, this raises questions of safety but in this context it suggests that Benveniste's agonist/

antagonist curve may correlate with clinical responses. We are already exploring this possibility of *in vitro* and *in vivo* overlap.

This landscape is so unfamiliar that perhaps we cannot see around our favourite biochemical landmarks to the horizon beyond. Doctors already observe the altered nuclear spin of diseased tissue in their patients with magnetic resonance imaging. Now Franco Bruno of Citta Universitaria Rome has communicated (OMHI Conference, Rome, April 1988) apparent confirmation of earlier work⁸ that the nuclear magnetic resonance spectra of homeopathic dilutions are altered. Here may be glimpses of the territory we seek.

DAVID TAYLOR REILLY

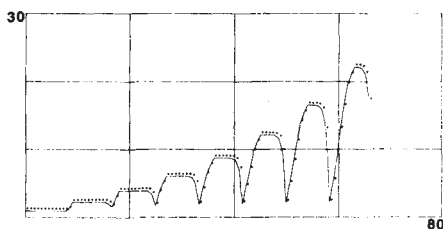
University of Glasgow,
University Department of Medicine,
Queen Elizabeth Building,
Royal Infirmary,
10 Alexandra Parade,
Glasgow G31 2ER, UK

1. Reilly, D.T., Taylor, M.A., McSharry, C. & Aitchison, T. *Lancet* **ii**, 881–886 (1986).
2. Boyd, W.E. *Br. Hom. J.* **32**, 106–111 (1942).
3. *L'Homeopathie en 1986 — une réalité sociale, scientifique, économique et industrielle* (Syndicat National de la Pharmacie Homeopathique, Paris, 1987).
4. Coulter, H.L. *Homeopathic Science and Modern Medicine* 69–70 (North Atlantic Books, California, 1980).
5. Blackley, C.H. *Br. Hom. J.* **29**, 238–286, 477–501, 713–736 (1871). *Ibid* **30**, 246–274, 417–449, 656–678 (1872). *Ibid* **31**, 77–103 (1873).
6. Millsbaugh, C.F. in *New Old and Forgotten Remedies* 16–17 (Boericke & Tafel, Philadelphia, 1900).
7. Poitevin, B., Davenas, E. & Benveniste, J. *Br. J. clin. Pharmac.* **25**, 439–444 (1988).
8. Young, T.M. *J. Am. Inst. Hom.* **68**, 8–16 (1975).

SIR—Davenas et al. (*Nature* **333**, 816–818; 1988) demonstrate that degranulation can be triggered at dilutions of anti-IgE antiserum in which not a single anti-IgE molecule is expected. In view of the revolutionary nature of this finding it is of the utmost importance to discuss all possible controls. As we feel an important control experiment has been overlooked, we would like to draw attention to a potential problem in the interpretation of the data of Davenas et al.

As described in the 'Methods' section of their article, Davenas et al. use microtitre plates to assay the basophil degranulation. Interestingly, these authors observe a periodicity in the basophil degranulation as a function of anti-IgE antiserum dilution. One might wonder to what extent this observation can be accounted for by contaminating, for example by aerosol, the contents of adjacent wells when filling the microtitre plate. As a contamination at the level of a single anti-IgE molecule might in principle corrupt the relevance of the findings, this criticism must be taken seriously.

In a microtitre plate, a contamination between adjacent wells will not result in a simple blurring of the measured titration profile. As the wells in a microtitre plate are placed in a matrix organization, any



Simulation of the filling of a conventional 8×12 well microtitre plate. The wells are filled sequentially from left to right. Upon the addition of a unit volume of anti-IgE antiserum (having a concentration which is halved at each step), 0.001 volume units are spread equally to neighbouring wells as described in the text. Abscissa: $-\ln(2^n)$, where n is the dilution step number. Ordinate: ratio of actual versus expected (i.e. in absence of contamination) concentration of anti-IgE.

contamination provoked by filling a given well will affect both its neighbours in adjacent columns as well as in adjacent rows. When filling the wells in a sequential order, wells in adjacent rows represent a jump by several orders of magnitude in the dilution series. Consequently, possible contamination will not merely result in a local blurring of the titration profile. To demonstrate that one expects oscillations to occur in the measured profile, we simulated the filling of a conventional 8×12 well microtitre plate. Using a simple model, in which upon filling a well at position (i,j) , a small fraction of its contents is spread out to its immediate neighbours having plate coordinates $(i \pm 1, j \pm 1)$, periodicities are observed in the concentration of effector molecules (anti IgE). This is illustrated in the figure which shows the ratio of actual to expected concentration of effector molecules at each dilution step using a 0.1 per cent contamination level.

Clearly, in view of this simulation, one cannot *a priori* rule out the possibility that the periodicity in degranulation observed by Davenas *et al.* is due to a contamination effect. In particular, our contamination argument questions whether the wells at highest dilution are really devoid of any anti-IgE molecule. We therefore suggest a control experiment in which microtitre plates are substituted with individual test tubes. Alternatively, one could interlace on a given microtitre plate the anti-IgE antiserum dilution and the control anti-IgE dilution series.

IGNACE LASTERS
MICHEL BARDIAUX

Plant Genetic Systems,
Laboratories Brussels—
ULB-UCMB-CP 160,
Avenue P. Heger, Batiment P2,
B1050 Brussels,
Belgium

SIR—Water has been the centre of many controversies during the past 20 years or so. In particular, scientists have repeatedly

proposed unusual properties for this 'universal' biological solvent. The case of 'super water' involving 'high tech' probing systems such as (15 years ago), nuclear magnetic resonancing was used to demonstrate that water could mediate special long-distance interactions. This has been proved to be due to a tricky artefact.

In the present case, I am puzzled by the fact that there has been no control of impurities (for instance by atomic absorption or neutron activation). In particular, the need for strong agitation (which is contradictory to the 'memory' hypothesis) suggests that test tubes might be involved. It seems worth recalling the experiments by Sternweis and Gilman, who have shown that the well-known fluoride effect no longer occurs when experiments are performed with pure water, in the absence of traces of aluminium¹. Indeed, they have shown that F^- is able to extract Al^{3+} from the glass surface, generating AlF_4^- which is responsible for the effects^{1,2}. Since it is well known that antibodies strongly (and often specifically) interact with surfaces, it is possible that they extract some ion (or other contaminant molecule), which in turn acts as a trigger for further extraction (in the absence of antibody). This would account for the requirement of strong agitation.

Proper control protocols should be performed before the generally efficient physicochemical laws are broken.

ANTOINE DANCHIN

Institut Pasteur,
28 Rue du Dr Roux,
75724 Paris Cedex 15, France

1. Sternweis, P.C. & Gilman, A.G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4888–4891 (1982).
2. Bigay, J. *et al.* *EMBO J.* **6**, 2907–2913 (1987).

SIR—In the leading article accompanying the article on an inexplicable observation about basophil degranulation by very dilute antiserum against IgE, you invite the vigilant reader to pick holes in the reported work. One obvious flaw can be seen when looking at the standard errors (s.e.) given in Table 1 of the article. According to the legend to Table 1 and to Fig. 1, the data represent the mean $\pm$ s.e. of basophil number actually counted in triplicate in a haemocytometer. The mean counts range from 27.7 to 106.7 and the s.e. are with two exceptions below 3 (2.6 per cent of the mean in average). However, according to the Poisson distribution, the s.e. of counting randomly distributed events, like cells in a haemocytometer, is equal to the square root of (mean counts/number of countings), that is, it ranges from 3 to 6 (11 per cent to 6 per cent of the means) for the triplicate counts given in Table 1. It is very unlikely that the observed s.e. of the triplicates are so much below the expected s.e. of counting.

The reason for this discrepancy is not clear. It is, however, unlikely that the s.e.

given in Table 1 are the result of an erroneous calculation, as the data of 4 triplicate experiments (Tyrode's-HSA) are actually displayed in Table 1, and again the s.e. of the experimental triplicates (lower than 1 per cent) are much below the s.e. expected already from counting 3×3 times 100 cells (3.3 per cent). The only way to explain the discrepancy would be to assume that actually more basophils have been counted and that the indicated numbers are somehow calculated values. However, in order to reach the average s.e. of 2.6 per cent as published in Table 1, at least 500 basophils would have to be counted in each sample.

In conclusion, the data in Table 1 are very unlikely to be derived from flawless experimental results and are certainly not the kind of data one would need to "throw away our intellectual heritage".

W. FIERZ

Universitätsspital Zürich,
Departement für Innere Medizin,
Haldeliweg 4,
CH-8044 Zürich,
Switzerland

SIR—I find it rather curious that a scientific journal should be so wary of non-material real phenomena as to refuse to endorse the findings of the Benveniste group research on the effect of dilution on anti-IgE antiserum! Does *Nature* expect nature to accommodate academic disciplines in order to be vindicated?

Ignorance about such non-material energetic effects as those discovered and demonstrably utilized in human medicine since the early nineteenth century by Carl Hahnemann and his successors or, more recently, those studied by Tesla, Morell and Popp in the field of physics, is hardly to the credit of a would-be authority on science; nor does it lend credence to *Nature's* claim of reliable and impartial reporting of significant new research. Casting doubt on findings merely because they are inconvenient to established assumptions and patterns of speculation strikes me as a poor way of advancing scientific knowledge.

KURT OPITZ

School of Maritime Studies,
Hamburg Polytechnic,
Rainvilleterrasse 4,
D-2000 Hamburg 50, FRG

SIR—The paper demonstrating that dilutions of anti-IgE must be vortexed rather than stirred in order to retain an imprint of the antibody on the solvent elucidates another long-standing question: how James Bond could distinguish Martinis that had been shaken or stirred.

A. NISONOFF

Rosenstiel Research Center,
Brandeis University,
Waltham, Massachusetts 02254-9110,
USA

"High-dilution" experiments a delusion

The now celebrated report by Dr J. Benveniste and colleagues elsewhere is found, by a visiting Nature team, to be an insubstantial basis for the claims made for them.

THE remarkable claims made in *Nature* (333, 816; 1988) by Dr Jacques Benveniste and his associates are based chiefly on an extensive series of experiments which are statistically ill-controlled, from which no substantial effort has been made to exclude systematic error, including observer bias, and whose interpretation has been clouded by the exclusion of measurements in conflict with the claim that anti-IgE at "high dilution" will degranulate basophils. The phenomenon described is not reproducible in the ordinary meaning of that word.

We conclude that there is no substantial basis for the claim that anti-IgE at high dilution (by factors as great as 10^{130}) retains its biological effectiveness, and that the hypothesis that water can be imprinted with the memory of past solutes is as unnecessary as it is fanciful.

We use the term "high dilution" reluctantly; these solutions contain no molecules of anti-IgE, and so are not solutions in the ordinary sense. "Solute-free solution" would similarly be illogical.

Our conclusion is based on a week-long visit to Dr Benveniste's laboratory, the INSERM unit for immunopharmacology and allergy (otherwise INSERM 200) at Clamart, in the western suburbs of Paris, during the week beginning 4 July. Among other things, we were dismayed to learn that the salaries of two of Dr Benveniste's coauthors of the published article are paid for under a contract between INSERM 200 and the French company Boiron et Cie., a supplier of pharmaceuticals and homoeopathic medicines, as were our hotel bills.

Benveniste's results are being widely interpreted as support for homoeopathic medicine. In the light of our investigation, we believe that such use amounts to misuse.

Our visit and investigation were preconditions for the publication of the original article. We acknowledge that we are an oddly constituted group. One of us (J.R.) is a professional magician (and also a MacArthur Foundation fellow) whose presence was originally thought desirable in case the remarkable results reported had been produced by trickery. Another of us (W.W.S.) has been chiefly concerned, during the past decade, in studies of errors and inconsistencies in the scientific literature and with the subject of misconduct in science. The third (J.M.) is a journalist with a background in theo-

retical physics. None of us has first-hand experience in the field of work at INSERM 200.

We acknowledge that we might well have found ourselves unable to get to grips with the work of the laboratory. But, on the basis of our experience, we are confident that the design of the experiments reported by INSERM 200 is inadequate as a basis for the claims made last month and that the defects we shall catalogue are a sufficient explanation of the remarkable results then reported.

We believe that experimental data have been uncritically assessed and their imperfections inadequately reported. We believe that the laboratory has fostered and then cherished a delusion about the interpretation of its data.

We are grateful to Dr Jacques Benveniste for his openness in discussing most of the questions we raised with him. He allowed us to borrow and to photocopy the relevant laboratory notebooks, which were invaluable for our investigation. We have every reason to believe that Dr Benveniste was (and, perhaps, still is) convinced of the reality of the phenomena reported in his article. We are also in the debt of several of Dr Benveniste's colleagues, especially to Dr Elisabeth Davenas. On her fell most of the burden of demonstrating the standard dilution experiments and of repeating them in a blinded protocol under our scrutiny. We know that our report will be a disappointment to the laboratory. We are sorry.

What follows is a narrative account of our visit and a summary of our conclusions.

Our investigations concentrated exclusively on the experimental system on which the publication was based. During our week in Paris, we resisted several proffered opportunities to examine other systems in which high dilution is claimed

not to diminish the biological effectiveness of a molecule.

The experimental system has evolved from a test for assessing the susceptibility of people to specific allergens. The guiding principle is that blood-borne allergens have the specific effect of interacting with the leukocytes known as basophils, causing them to degranulate — that is, to release the contents of cytoplasmic granules carrying histamine and other active substances provoking the symptoms of asthma and hay-fever.

These allergic reactions are apparently mediated at least in part by IgE molecules attached to the surfaces of basophils (in the blood) or mast cells (in tissues). Normally, degranulation is triggered by the interaction of anchored IgE molecules with an antigen, but the same effect can be brought about by the use of anti-IgE — antibody prepared by injecting human IgE into an animal of another species. (INSERM 200 uses goat anti-IgE at a concentration of 1 mg cm^{-3} sold by the Dutch company Nordic.)

The laboratory notebooks provide ample evidence that this expected degranulation is a maximum between $\log(\text{dilution})$ 2 and 4.

Benveniste described the published procedure as a "simple experiment". A buffered solution of anti-IgE is serially diluted by a factor of 10 by transferring measured volumes from one test-tube to another. Pipette tips are discarded after each transfer. Measured volumes of re-suspended white cells derived from human blood are transferred to wells in a polystyrene plate. To each of these is added a measured volume of serially diluted anti-IgE or buffer as a control. The wells are incubated for 30 minutes at 37°C . An acidic solution of toluidine blue, which stains intact but not degranulated basophils red, is added and the numbers of recognizable basophils counted on a haemocytometer slide. Anti-IgG, which does not degranulate basophils, is used as a control.

We were surprised to learn that the experiments do not always "work". There have been periods of several months at a time during which solutions at high dilution have not degranulated basophils. Indeed, the laboratory had just emerged from such a period. (Speculation at the laboratory is that the distilled water may have been contaminated, or otherwise made unsuitable.) It also appears that

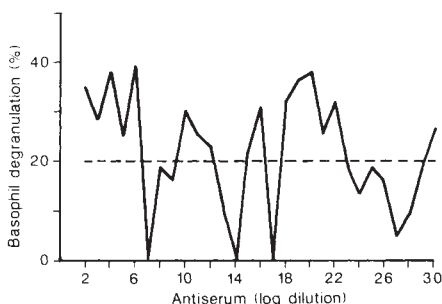


Fig. 1 A demonstration degranulation, the first of the three open experiments.

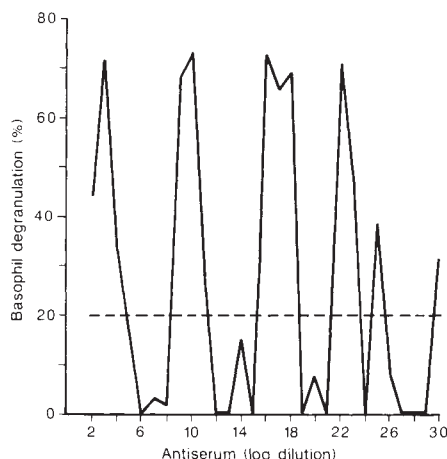


Fig 2 The fourth demonstration experiment (read "blind") with unexpectedly high peaks (see text).

bloods that "do not degranulate" are often encountered; we were informed that, in this event, data are recorded but not included in analyses prepared for publication. Even so, the source of blood for the experiments is not controlled, except that an attempt is made not to use blood from people with an allergy.

We witnessed a total of seven runs of this experiment, of which three were routine repetitions of the standard procedure. For the fourth experiment, samples of diluted IgE were transferred by one of us (W.W.S.) to wells in a plastic plate in a random sequence and then read blind by Dr Davenas. All four of these experiments, the last after decoding, gave results described as positive by Benveniste. But three further sequences of counts of stained basophils in three further strictly blind experiments gave negative results (see below).

Figure 1 shows results gathered in the first group of experiments. The ordinate is the decrease (compared with the control) of the numbers of stained basophils at dilutions ranging from 10^{-2} to 10^{-30} . In each case, the left-most peak is that expected from the interaction between anti-IgE and IgE bound to basophils. The number of stained basophils increases to near its control value at log(dilution) of between 5 and 7 (0 per cent degranulation, called "achromasie"); the unexpected phenomenon is that the graphs then reach a series of three or four further peaks with increasing dilution.

These are the successive peaks of activity said in the original article to occur in a periodic fashion, and whose position was said to be reproducible. It is clear from the four graphs that this claim is not obviously supported by this data. The laboratory notebooks confirm that the position of the peaks varies from one experiment to another.

The data in the fourth experiment appear different from those recorded earlier in the laboratory. Indeed, Ben-

veniste volunteered that "we've not seen one like this before". The odd feature of the curve is that the activity of the diluted anti-IgE is, at its peak, identical with that of anti-IgE at log(dilution) 3 — presumably the point at which the natural degranulating effect of anti-IgE is a maximum.

We raised with Dr Benveniste and his colleagues the obviously relevant question of the sampling error. We were astonished to learn, in the discussion of our conclusions at the end of our visit, that neither Dr Benveniste nor his colleagues seemed to be aware of what sampling errors are. We provided a simple explanation, complete with an account of what happens when one pulls a handful of differently coloured balls from a bag, to argue that the sampling error of any counting measurement must be of the order of the square root of the number to be counted. On several occasions, Benveniste called these "theoretical objections".

Ironically, he is himself one of the three authors of a paper published in 1981, in which just this issue had been addressed in a superficially similar situation (Petoit, J.F., Sainte-Laudy, J. & Benveniste, J. *Ann. Biol. clin.* **39**, 355; 1981), and which appears to be the justification of the dotted line drawn at about 20 per cent (corresponding to two standard deviations) on the per cent degranulations of intact basophils after axis.

That brief paper deals exclusively with the effect of sampling errors (not other kinds of errors) on the interpretation of measurements of intact basophils after white-cell suspensions had been allowed to react with allergens via their attached IgE molecules. Even now, at the Clamart laboratory, provision is made for the measurement of two control samples. Among other things, the paper provides a statistical test for telling when the difference between the two control values is statistically significant at the 5 per cent level, in which case people using the procedure as a diagnostic test of allergy are advised to start their experiment all over again.

At INSERM 200, there seems to have grown up a less formal way of dealing with problems of this kind; when the reading of a diluted sample is greater than the control counts, the experimenter often counts the control sample again, on the grounds that the first reading "must have been wrong". This happened when Dr Davenas was counting the first of the first group of experiments.

This procedure exaggerates to some extent the amount of basophil degranulation measured with reagents at high dilution. The practice makes the control values unreliable, and is a significant pointer to the laboratory's disregard of statistical principles.

In these circumstances, it is natural that

we should eagerly have accepted Benveniste's invitation to devise a blind experiment. We set out to devise a procedure that would be watertight. We asked that three samples of blood should be run. The serial dilutions would be prepared by Dr Davenas, secretly coded by us before being transferred to wells for incubation and staining by her.

In a small laboratory, procedures like this are inevitably and understandably disruptive. At INSERM 200, the sense of melodrama was further heightened by the general recognition of the importance of the trial, and by the precautions necessary to ensure that the code would not be known to others than ourselves as well as by the need that the one of us with a reputation for sleight of hand (J.R.) could be shown to have been kept away from the test-tubes containing the serial dilutions.

This was done by arranging that Davenas should carry the diluted anti-IgE solutions in stoppered test-tubes to a separate room, where their contents would have been transferred to previously labelled tubes as determined by counters drawn at random from a bag. The coding procedure was monitored by a video camera operated by Randi, who was thereby prevented from touching anything else. (We have a record of the proceedings on an unbroken reel of tape.)

We made two last-minute changes in the planned procedure. First, we included 5 control tubes containing only buffer.

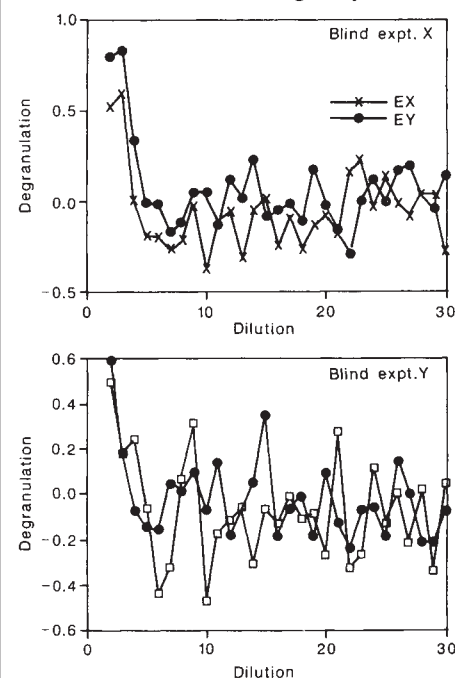


Fig. 3 Records for the first two blind experiments (5–7 inclusive), showing sampling noise only below the expected decline of degranulation with increasing dilution. Note that the ordinate extends below zero on the degranulation scale (to accommodate sampling errors above as well as below the control values).

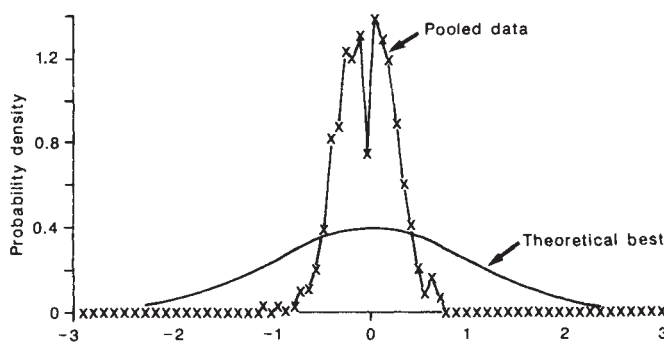


Fig. 4 Comparison of measured departures of duplicate normalized readings from their means with the gaussian distribution expected.

Second, having been warned that homoeopaths might regard the data as invalid if solutions were decanted from one set of tubes to another, we removed the numbers written with a felt pen on the original tubes, replacing them with numbered labels which Randi assured us were tamper-proof. The code itself was eventually folded in aluminium foil, enclosed in an envelope specially sealed by Randi and then taped to the laboratory ceiling for the duration of the experiment.

We also arranged a second step of coding just before the slides were counted. One of (W.W.S.) took responsibility for pipetting, after securing the agreement of Davenas and Benveniste that his technique was satisfactory. Both the laboratory procedures and the codes themselves were recorded on video tape. The plates containing the stained cell suspensions were stored in a box (sealed by Randi) in a cold room until read, in random order. The second plate took longer to read, partly because each well was read in duplicate by each observer (Dr Davenas and her colleague, Dr Francis Beauvais), partly because the cells of the second plate were only faintly stained and were thus difficult to read.

Whatever the three runs would provide, we were especially anxious to derive some objective estimate of the intra- and inter-observer measurement errors. We had been told at the outset, by Benveniste, that Dr Davenas was not merely exceptionally devoted to her work but the one in whose hands the experiment most often "works". He said that she usually "counts more cells" than other people. Dr Beauvais, who was also said to be exceptionally skilled, read the slides separately from Dr Davenas, but at the same time. On this occasion, the sampling errors missing from most of the laboratory records did indeed appear.

The duplicate measurements in our strictly blinded experiments were especially important. First, they show that sampling errors do indeed exist, and are not "theoretical objections". Second, they show that the two observers were counting as accurately as could be expected, which gives the lie to the later complaint that the

results of the double-blind experiments might be unreliable because the observers had been exhausted by our demands.

Others working in this field recognize the difficulty of counting basophils (roughly 1 in 100 among leukocytes), preferring instead to measure the histamine released on degranulation. This practice is not followed at INSERM 200 because, we were told, of previous failure to record histamine release (as distinct from the disappearance of stained basophils) at high dilution (whence the term "achromasie").

We began to break the codes by lunchtime on our last day, the Friday. When the slides had been matched to the wells from which their samples had derived, but before the appropriate dilutions had been assigned to them, there was a great sense of light-heartedness in the laboratory, no doubt at the prospect that the ordeal would soon be at an end. Benveniste, glancing at the half-decoded data, even offered to predict where the peaks and troughs would fall in the data. His offer was accepted. But his predictions proved to be entirely wrong.

We asked at this stage for criticisms of the conduct of the trials, but were given none. To the question what would be said if the two observers had recorded degranulation peaks, but at different high-dilution values, Benveniste said that would still constitute success.

Opening sealed envelopes is Randi's expertise. He found that the sealed flap of the envelope had detached itself at a surprisingly straight angle when the scotch tape attaching the code to the ceiling was

pulled away, but inspection of the aluminium foil allowed him to pronounce himself satisfied that the code had not been read. Then came the decoding — one person singing out numbers to another.

So do the numbers make sense? Six numbers into the record of the first plate to be read, Benveniste said "that patient isn't degranulating, try another". So we did — first the parallel readings by Dr Beauvais, then the remaining two experiments. In the event, the results of all three experiments were similar. The anti-IgE at conventional dilutions caused degranulation, but at "high dilution" there was no effect. Blood from three sources in a row degranulated at ordinary dilutions but not at homoeopathic dilutions. Each of the three experiments was a failure.

Conclusions

We conclude that the claims made by Davenas *et al.* are not to be believed. Our conclusion, not based solely on the circumstance that the only strictly double-blind experiments we had witnessed proved to be failures, may be summarized as follows:

■ The care with which the experiments reported have been carried out does not match the extraordinary character of the claims made in their interpretation. What we found, at Clamart, was a laboratory procedure possibly suitable for the application of a well-tested bioassay, but unsuitable as a basis for claiming that anti-IgE retains its biological activity even at a log(dilution) of 120. In circumstances in which the avoidance of contamination would seem crucial, no thought seemed to have been given to the possibility of contamination by misplaced test-tube stoppers, the contamination of unintended wells during the pipetting process and general laboratory contamination (the experiments we witnessed were carried out at an open bench). We have no idea what would be the effect on basophil degranulation of the organic solvents and adhesives backing the scotch tape used to seal the polystyrene wells overnight, but neither does the laboratory.

The design of the experiments hardly matches the nature of their interpretation.

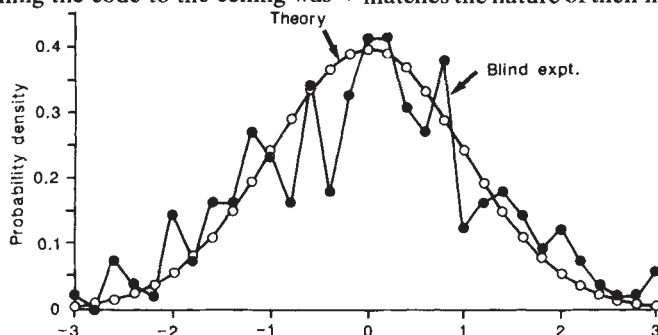


Fig. 5 Same as Fig. 4 except that data derive from duplicated readings within the blind experiments only.

For example, one would have thought that counting wells at least in duplicate would have been an elementary precaution against gross errors. The second of our strictly blinded experiments seems to be one of the few in which something of this kind had been attempted.

The laboratory seems to have been curiously uncritical of the reasons why its experiments do not, on many occasions, "work". For example, we were told that the best results were obtained when cells were left in the cold-room overnight before counting, but there has been no investigation of that phenomenon, or of the reports that taking a second sample from a single well gives odd results (an effect not apparent in our double-blind experiments).

■ **The phenomena described are not reproducible, but there has been no serious investigation of the reasons.** We have referred to the fact that some blood yields negative results, and that there are periods of time when no experiments work. But the laboratory notebooks show great variability in the positions at which peaks occur.

■ **The data lack errors of the magnitude that would be expected, and which are unavoidable.** This is best illustrated by Fig. 3, whose two graphs have been constructed from data recorded by Dr Davenas from samples supposedly identical with each other, usually measurements of control samples but also including some duplicate runs. The recorded values have been normalized by subtracting the mean and dividing by the square root of the mean (the expected sampling error). If the only source of error were sampling error, the standard deviation of the plotted curve should be unity (1). Other sources of error, for example, experimental variability, could only increase the standard deviation. But Fig. 3 shows that repeat observations agree more closely than would be expected from the underlying distribution. This is a well-known effect that sometimes affects duplicate readings by the same individual, but the magnitude of the effect in this case calls into question the validity of the readings. This artefact is nevertheless not apparent in the blinded duplicated readings.

■ **No serious attempt has been made to eliminate systematic errors, including observer bias.** It is true that the laboratory notebooks record experiments in which anti-IgG has been used as a control; we were surprised to find that the IgG control run reported by Davenas *et al.* (their Fig. 1b) was carried out at a different time from the run with IgE published in the same figure.

Most of the data recorded in the laboratory notebooks derive from experiments in which the same person has been responsible for the sequential dilution, plating out and counting. Given the

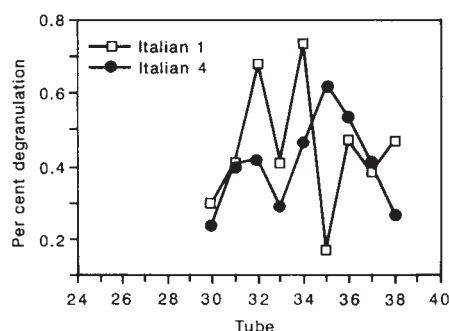


Fig. 6 Two duplicate Italian runs showing high degranulation, but discordantly.

shared belief at Clamart in the reality of the phenomenon reported last month, and its potential importance, it is mystifying that duplicate and blind counting is not routine.

■ **The climate of the laboratory is inimical to an objective evaluation of the exceptional data.** So much is readily apparent from the way in which experiments are described as successes and failures, by the use of the word "working" to describe experiments yielding a positive result, and by the several speculations we were offered, without experimental evidence in their support, to explain the several failures the laboratory has experienced. The folklore of high-dilution work pervades the laboratory, as epitomized by the suggestion that decanting diluted solution from one tube to another might spoil the effect and the report that the repeated serial dilution by factors of three and seven (rather than ten) always yields negative results.

Collaborations

We have not been able to pay as much attention as we would have wished to the data collected at other laboratories and cited in Davenas *et al.*, but we have examined documentary evidence available.

Supporting data were said to have come from Rehovot (Israel), Milan and Toronto. Dr Benveniste told us we could not see the Toronto data, described as preliminary, without the consent of the authors, who could not be telephoned.

The data gathered in Israel and Milan are, apparently, significant. Figure 6 is typical of the data from Milan. Though there are no duplicate measurements and therefore no direct evidence of sampling error, there is also some evidence of degranulation at high dilution. Without knowing more about the circumstances, we are unable to comment.

The Israeli data are more extensive. The first trials were in March 1987, during a visit to Rehovot by Dr Davenas. The most remarkable of several successful trials was her correct identification of seven high-dilution tubes out of ten presented to her blind. Even so, the report (to Benveniste) of the trials was cautious. Later, analysis of the tubes which had

tested positive in this trial revealed not merely immunoglobins but other protein contaminants apparently identical with materials in the original IgE vial. One of the participants (Professor Meir Shinitzky of the Weizmann Institute) then withdrew as a putative co-author.

Since then, there have been two developments in Israel — a series of experiments carried out independently of Benveniste's laboratory and a further blinded experiment. Data from the latter are unfortunately not available. Maître Simart, a legal official at Clamart who held the codes, is said not to have had time to decode them.

These measurements are nevertheless, to judge from the documents we have seen, stronger evidence than any we found at Dr Benveniste's laboratory to support his claims. But we do not have the information to evaluate them.

Postscript

We presented the substance of these conclusions to Dr Benveniste and his colleagues immediately after the strictly blinded experiments were decoded. The discussion that followed was inevitably tense. Benveniste acknowledged that his experimental design may not have been "perfect", but insisted (not for the first time) that the quality of his data was no worse than that of many papers published in *Nature* and other such journals.

One of us (J.M.) said it would be best if Benveniste would withdraw the published article, or at least write to *Nature* to qualify his findings and their interpretation, in which case we would not publish this report. It was mutually agreed that nothing would be said publicly until 28 July. But Benveniste said that the laboratory would work through the weekend "and all next week" to prove the reality of the phenomenon.

Our greatest surprise (and disappointment) is that INSERM 200 seems not to have appreciated that its sensational claims could be sustained only by data of exceptional quality. Randi put the point best, during our Friday discussion, by saying: "Look, if I told you that I keep a goat in the backyard of my house in Florida, and if you happened to have a man nearby, you might ask him to look over my garden fence, when he'd say 'That man keeps a goat'. But what would you do if I said, 'I keep a unicorn in my backyard'?" We have no way of knowing whether the point was taken.

Eventually, there was no more to say. We shook hands all round, sped past the common-room filled with champagne bottles destined now not to be opened and into the lens of a news agency photographer summoned for the happier event.

John Maddox
James Randi
Walter W. Stewart

Dr Jacques Benveniste replies:

AMAZINGLY, J. Maddox, with all his experience, fell with us into the trap set by a squad of "self-appointed keepers of the scientific conscience", "with no substantial scientific published record" (J. Maddox, *Nature* 333, 795; 1988). Their amateurism, the climate they created in the five days of our ordeal, their inability to get to grips with our biological system and their judgement based on *one* dilution series dismiss this inquiry altogether. Who, with even the slightest research background, would blot out five years of our work and that of five other laboratories on such grounds?

For two years, I asked *Nature* to check our data. But the magician and the invigilator defined above worried me deeply. Mr Maddox assured me that he would prevent any wrongdoing. In fact, a tornado of intense and constant suspicion, fear and psychological and intellectual pressure unfit for scientific work swept our lab. Furthermore, these lesson-givers were astonishingly incompetent. In spite of my demands, no programme was set beforehand.

There were performed in 5 days 3×30 ten-fold dilutions, preparation and degranulation (35 tubes each) of 7 leukocyte samples, and eye scan of 300 chambers (about 20,000 basophils). Half of that is way beyond the weekly individual limit. The first two days of the week were spent on four open experiments. The first blood did not react even to high anti-IgE, but the three other results were superb. The fourth (counted blind upon our insistence) was "incredible": 70–75 per cent degranulation at dilutions 10, 16/18, 22, similar to Fig. 1b of the article, controls varying by the usual 15.

Then Stewart, with his typical know-it-all attitude, called these results, blind though they were, valueless; that implies fraud before counting. The third day, a new dilution series was *single-coded* in front of a video camera, involving two major professional errors since all the visitors knew the code, when "to believe the unbelievable"? (The witness camera could not record time, nullifying that part of the procedure.) The code, wrapped into aluminium foil and in an envelope, was taped to the ceiling!

The next day, the hysteria was such that Maddox and I had to ask Stewart not to scream. He had decided also to blind the counting (an overkill) and to fill the chambers, using a modified untested method (two other serious errors). Referees must respect experimental design and not take part in it. This one was untrained

and knew both codes (dilution and counts).

Here is another hard-to-believe incident: Stewart imposed a deadly silence in the counting room, yet loud laughter was heard where he was filling chambers. There, during this critical process, was Randi playing tricks, distracting the technician in charge of its supervision!

It will now be clear what a mockery of scientific inquiry this was. Only the constant implication that we had something to hide (the squad left with 1,500 photocopies!) prevented me from stopping this masquerade. On one blood, basophils could barely be counted. On the two others, controls ranged from 40 to 81 for one operator, from 35 to 61 for the other, the worst ever.

Duplicates such as 39–63 were found; if 39 were right, degranulation would have been 61 per cent at dilution 22. Thus, the first three open and blind tests worked, controls being impeccable, whereas on the last days the test worked poorly mainly due to erratic controls. Something happened, probably the work load and modifications enforced by the "expert".

All in all, the judgement is based on one dilution tested on two bloods in awful technical and psychological conditions. Outrageous! Then, the team flew away in minutes, not leaving behind any report, nor even the data that I had to collect at Stewart's hotel that night! The report is filled with inaccuracies and distortions. Just a few: does the fact that homeopathic companies are paying two researchers (contract approved by INSERM administration) mean that they order them into improper conduct? How about research in — or supported by — industry, including numerous Nobel prizes? We could not self-finance a long-term international cooperation nor the expenses of this large group of investigators. Did the source of the money influence their judgement? What a level of argument!

The Scotch tape was placed above all wells (see controls below). Repeating a wrong count? This detects counting errors, especially when stressed by pointed microphones and camera. The two closest counts are chosen at risk of being all wrong, as one recalibrates an ill-tuned machine, even in a blind experiment. The central argument bears on sampling errors and statistics of which we are so aware that we performed numerous control experiments. They show similar standard deviations and variances in 24/28 comparisons of blind (4 series, 90 samples, without the Israeli experiments) versus

open (7 series, 183 samples) control wells.

Did the "experts" understand that the real controls are water or anti-IgG most often paired with anti-IgE (Fig. 1b)? They analysed a few curves out of 1,500 pages, but most positive data are anyhow way off 1 or 2 standard deviations. Other allergy tests correlate with degranulation (reference in article), so why is it that our statistics fit for 40 to 70 per cent degranulation at regular ligand concentration and not for the same at high dilution?

Similar double-blind experiments (*Br. J. clin. Pharm.*) were under the control of an INSERM statistician, using a better non-parametric test seemingly unknown to our visitors. Then, the report auto destroys the statistical bias, declaring it "not applicable to all" (how many?) "data, for example in the 4th experiment" similar to Fig. 1b or to the double-blind tests supervised by our Dean and a bailiff or Israeli scientists (tables).

Being statistically sound (which is "bloody obvious" using common sense), are all these results "made up" as snapped at me by Stewart, the very referee who cleared the paper with raw data and statistics in hand? Why then accept a paper on 13 June to publish June 30th to destroy on 8 July data so easily spotted as wrong or made up? Is it a display to the world of the almighty anti-fraud and heterodoxy squad? Lip service is paid to our honesty; yet accusation of cheating was rampant, as shown by dismissal of the 4th experiment, Randi's mere presence and his lengthy examination of the supposedly violated code. This impinges on our honesty and scientific ability but also, without examination, on the other participating laboratories, which is unacceptable. I welcome academic exchanges on errors, if any, but will no more stand suspecting us or our associates.

More, I now believe this kind of inquiry must immediately be stopped throughout the world. Salem witchhunts or McCarthy-like prosecutions will kill science. Science flourishes only in freedom. We must not let, at any price, fear, blackmail, anonymous accusation, libel and deceit nest in our labs. Our colleagues are overwhelmingly utmost decent people, not criminals. To them, I say: never, but never, let anything like this happen — never let these people get in your lab. The only way definitively to establish conflicting results is to reproduce them. It may be that all of us are wrong in good faith. This is no crime but science as usual and only the future knows. □

J. Benveniste is at INSERM U200, Université Paris-Sud, 32 rue des Carnet, 92140, Clamart, France.

RNA translation

Picornaviruses break the rules

Richard J. Jackson

THE mechanism of translation initiation on picornavirus RNAs has always been an intriguing problem. These RNAs have unusual properties which hint that, unlike almost all other eukaryotic messenger (m)RNAs, they may not conform to the scanning ribosome model developed by Marilyn Kozak¹⁻³. The cardinal rule of this model is that ribosomes can access initiation sites only by first binding to the 5' end, followed by linear scanning of the mRNA; commitment to initiation is usually at the first (5'-proximal) AUG codon¹, although selection of more distal AUG codons is permitted under certain defined circumstances^{2,3}. What the model specifically excludes (even on uncapped mRNAs) is any type of direct internal initiation process which bypasses the 5' end³. But now, in an exciting development, Pelletier and Sonenberg on page 320 of this issue⁴ claim that there is efficient internal initiation on poliovirus RNA, and Jang *et al.* in a paper in the *Journal of Virology*⁵ make the same claim for encephalomyocarditis (EMC) RNA.

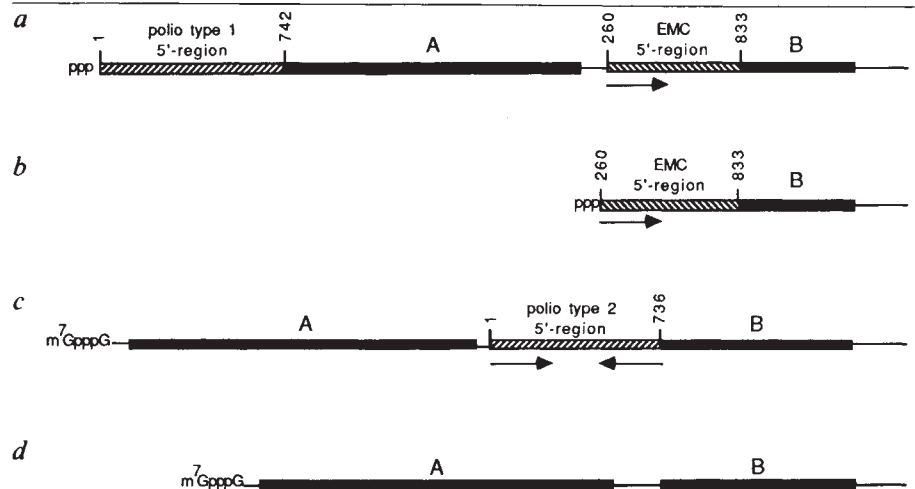
One peculiarity of picornavirus RNAs is the small protein (VPg) covalently linked to the 5' end of the virion RNA. As this is rapidly removed in cell-free extracts and *in vivo*, however, the RNAs are actually translated in the uncapped form. Because uncapped mRNAs are usually inefficient in comparison with capped, it is very surprising that EMC RNA and foot-and-mouth disease virus (FMDV) RNA are two of the most efficiently translated mRNAs known, and that the addition of a 5' cap to poliovirus RNA does not influence its translatability⁶.

All picornavirus RNAs have one very long open reading frame starting far from the 5' end of the nucleotide (N) sequence, at N 743 in type I poliovirus RNA, N 834 in EMC RNA, and even more distant in FMDV RNA. In all viruses, the long 5'-untranslated region has several AUG codons and short open reading frames, but no translation products have been detected, nor are these reading frames as strongly conserved between different serotypes of the same virus as would be expected if they encoded functional proteins. According to the scanning ribosome model, at least some of these short open reading frames should be efficiently translated^{1,2}, and such translation could quite severely reduce the initiation frequency at downstream AUG codons (including the main initiation site), certainly if the reading frames overlap^{3,7,8}. This problem is aptly illustrated by EMC RNA, where translation starts with

exceptional efficiency at the underlined AUG (N 834) of the published sequence⁹ . . .ACGAUGAUUAUAUGGCC., yet the model clearly predicts initiation exclusively at the nearby upstream (out-of-frame) AUG.

The recent evidence for internal initiation comes from experiments exploiting the fact that translation of a dicistronic mRNA with two non-overlapping reading

that the synthesis of B observed with dicistronic mRNA cannot be explained as initiation (by the standard scanning mechanism) on monocistronic mRNA fragments generated by degradation. Two results of Jang *et al.* surely overrule this objection: first, the yield of B from the full-length dicistronic construct is as high as from any monocistronic mRNA derivative tested; and second, when various deletions are made in the EMC segment, the synthesis of B from the dicistronic mRNA requires the whole EMC segment, whereas synthesis from the monocistronic mRNA does not. Moreover, this group has preliminary evidence that the EMC sequence also allows internal initiation on



Schematic diagram of the mRNAs discussed in the text. *a* and *b* are the dicistronic mRNA and the monocistronic control, respectively, studied by Jang *et al.*⁵. A is the viral oncogene *sea* and B is the coding sequence for poliovirus polypeptide 2A. *c* and *d* are the experimental and control dicistronic constructs, respectively, studied by Pelletier and Sonenberg⁴, where A is the herpes simplex virus-1 thymidine kinase gene and B is the chloramphenicol acetyltransferase gene. Coding sequences are depicted by black boxes and picornaviral 5'-untranslated sequences by cross-hatching, using the numbering system of the virion RNA nucleotide sequence. The small open reading frames in these viral sequences are not shown. Horizontal arrows depict the direction of deletions tested.

frames (A and B) generally gives a low yield of protein B compared with A^{3,7,8}. This result is consistent with the scanning ribosome model, which also predicts that B synthesis should be delayed relative to A, as the initiation site of B can be used only by those ribosomes that resume scanning after first translating cistron A (ref. 3). The aim, therefore, was to test whether placing the 5'-untranslated segment of a picornavirus genome between the two reading frames allows the downstream cistron to be translated independently of the upstream (see figure).

By using a 574-nucleotide segment (N 260-833) of EMC RNA, Jang *et al.*⁵ obtain the striking result that the rabbit reticulocyte lysate system synthesizes protein B earlier and in much higher yield than A. Indeed, the yield of B from this dicistronic mRNA is as high as from a monocistronic mRNA with the same upstream EMC sequence. As no *in vitro* translation assay has yet been invented that does not degrade the input RNA at least to a limited extent, we need to ensure

di- and tricistronic mRNAs in intact cells.

The same type of *in vivo* experiment is reported by Pelletier and Sonenberg in this issue⁴. They use a construct in which the entire 5'-untranslated region (736 nucleotides) of type 2 poliovirus is placed in the intercistronic region of a capped dicistronic mRNA (Jang *et al.* study uncapped mRNA). When the cells expressing the dicistronic mRNA are infected with poliovirus, the synthesis of protein A is inhibited and B enhanced, demonstrating that downstream cistron translation is independent of upstream. In addition, cell-free extracts from poliovirus-infected cells translate cistron B but not A. (Poliovirus infection of cells is known to shut off translation of all capped mRNAs, apparently because it inactivates eIF-4F, an initiation factor required for translation of capped, but not uncapped, mRNAs.) Fragmentation of the mRNA to yield monocistronic B mRNA cannot be the explanation for this selective synthesis of B, as the RNA actually undergoing translation *in vivo* and *in vitro* has been shown

to be intact dicistronic mRNA.

Both reports are incompatible with a mechanism that requires linear scanning from a 5' end, assuming that the far-fetched possibility of a type of nicking-closing activity generating a transient 5' end within the picornaviral segment can be rejected. It also seems unlikely that ribosomes bind first to the 5' end of the dicistronic mRNA and then make a long-range 'jump' dictated by the viral sequences, although it could be argued that this can be rigorously excluded only by repeating the experiments with closed circular RNA constructs. The most reasonable explanation for the results, favoured by both groups^{4,5}, is that ribosomes bind directly to some internal site within the picornaviral segment, not only on the artificial dicistronic mRNAs but also when translating the viral RNA itself. On present evidence, this internal initiation requires as much as about 500 nucleotides of the viral 5'-untranslated segment^{4,5}, excluding the extreme 5'-proximal sequences but including most of the short open reading frames.

This surprising length leads Pelletier and Sonenberg¹ to suggest that the initiation site of cistron B could be selected by ribosomes scanning the RNA after first binding at some distance upstream, possibly at a region sufficiently unstructured to allow the initiation of mRNA unwinding believed to be essential for the scanning process. The attraction of this hypothesis is that it invokes a mechanism similar to the conventional scanning model, but it shares the difficulty of all such models in explaining how the scanning handles the many short open reading frames and upstream AUG codons in this viral sequence.

The alternative explanation, that ribosomes bind directly to the AUG codon of cistron B (or very near it), cannot account for the extreme length of viral RNA necessary for internal initiation. But it is possible that the function of this long viral RNA segment is to provide a complex higher-order structure which somehow causes ribosomes to bind directly to the main initiation site. The resolution of these questions will require more detailed definition of the minimal RNA sequences necessary, and a comparison of the initiation factor requirements of internal initiation as opposed to initiation site selection by scanning from the 5' end.

Virus-coded polypeptides do not seem to be required for this internal initiation: Jang *et al.* use an uninfected cell extract and observe that initiation of cistron B translation starts immediately, before any hypothetical products of translation of the short EMC reading frames could have accumulated to a significant extent. This implies that the machinery for internal initiation is present in uninfected cells and potentially operative on cellular mRNAs.

Obviously, the translation of most cellular mRNAs conforms to the strict rule of exclusive ribosome entry at the 5' end^{1,3}, but the possibility of rare cases of internal initiation raises exciting prospects. In addition, these results have far-reaching biotechnological implications for the problem of how to maximize the expression of a cloned gene in mammalian cells when it is not possible to devise a selection procedure for such expression. This difficulty can now be overcome by making a dicistronic construct comprising the

desired (unselectable) gene, the critical picornavirus sequences and a selectable marker gene. □

1. Kozak, M. *Microbiol. Rev.* **47**, 1-45 (1983).
2. Kozak, M. *Cell* **44**, 283-292 (1986).
3. Kozak, M. *Molec. cell. Biol.* **7**, 3438-3445 (1987).
4. Pelletier, J. & Sonenberg, N. *Nature* **334**, 320-325 (1988).
5. Jang, S.K. *et al. J. Virol.* (in the press).
6. Pelletier, J. *et al. Molec. cell. Biol.* **8**, 1103-1112 (1988).
7. Peabody, D.S. *et al. Molec. cell Biol.* **6**, 2704-2711 (1986).
8. Kaufman, R.J. *et al. EMBO J.* **6**, 187-193 (1987).
9. Palmenberg, A.C. *et al. Nucleic Acids Res.* **12**, 2969 (1984).

Richard J. Jackson is in the Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK.

Atomic physics

Ions that fit into place

Peter Knight and Richard Thompson

PLASMAS trapped in external electromagnetic fields can be cooled using laser techniques so that they eventually 'crystallize' into a regular three-dimensional array. Such crystals are tenuous, fragile objects that have lattice spacings up to many thousand times larger than those of normal solid-state crystals. H. Walther and co-workers from the Max-Planck-Institut für Quantenoptik report elsewhere in this issue (Blümel, R. *et al. Nature* **334**, 309-313; 1988) the direct imaging of the transitions between the crystalline and disordered states with relatively few ions. Significantly, there is hysteresis in the transition, ordering taking place at different conditions from the disordering.

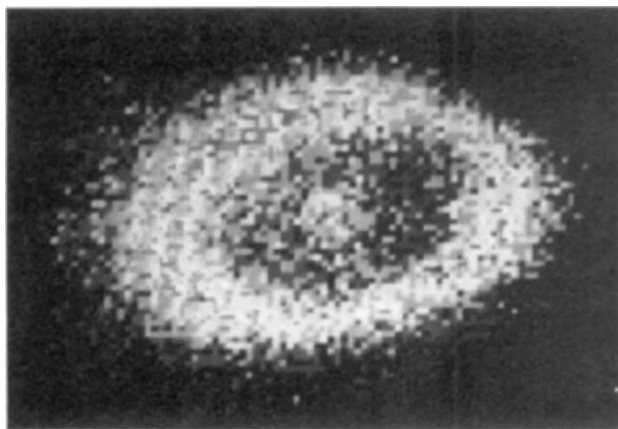
In the new experiments, magnesium ions are trapped in a radio-frequency 'Paul' trap. They are cooled by momentum-transfer processes in which many cycles of stimulated absorption of light, from a laser tuned just below a resonant frequency, and isotropic spontaneous emission reduce the temperature below 100 mK in a few seconds. The ion motion

can be viewed by detecting the fluorescence with a photon-counting image system. When the trap is first loaded with several ions, the random motion of the uncorrelated ions produces a cloud-like structure in the fluorescent image. As the laser cooling progresses, the kinetic temperature drops below the critical value which allows crystallization.

This critical temperature is governed by the localizing trap potential. A trapped ion cloud is really a one-component plasma comprising a single charged species embedded in a background trapping field. Coulomb repulsion between the ions competes with the confining potential of the trap. At high temperatures the ions have sufficient kinetic energy to move almost independently. Ordering occurs in two stages when the thermal energy is reduced below the nearest-neighbour Coulomb energy of repulsion. The first should occur when the thermal energy is less than half the Coulomb energy and short-range order as found in liquids arises. The second, 'liquid-solid' transition occurs after a further 200-fold

decrease in thermal energy.

Using a 'gas' of charged dust particles, Wuerker and co-workers first saw such crystallization and melting 30 years ago (Wuerker, R.F., Shelton, H. & Langmuir, R.V. *J. appl. Phys.* **30**, 342-349; 1958). Walther and co-workers observed the crystallization of 2-50 ions last year, as did Wineland and co-workers at the National Bureau of Standards, Boulder (Diedrich, F. *et al. Phys. Rev. Lett.* **59**, 2931-2934, and Wineland, D.J. *et al. Phys.*



This two-ion crystal was observed for 3 minutes. The cooling laser was set slightly off-axis, causing the crystal to rotate; hence the elliptical 'rim' with semi-major axes of 24 and 17 micrometres. Towards the end of the observation, the crystal 'exploded', one ion leaving the trap, the other eventually cooling to the centre; hence the 'hub' of the wheel. (Courtesy of H. Walther.)

Rev. Lett. **59**, 2935–2938; 1987). In a recent experiment using laser-cooled beryllium ions in a Penning trap, which employs static electric and strong magnetic fields, the American group has studied transitions between disordered cloud motion, ordered shell structures with diffusion only within each shell and, finally, a fully crystallized state (Gilbert, S.L., Bollinger, J.J. & Wineland, D.J. *Phys. Rev. Lett.* **60**, 2022–2025; 1988).

The new work by Walther and co-workers, reported in this issue, concentrates on the nature of the phase transition between the cloud and crystal forms of the laser-cooled ions. As they vary the laser power, the detuning of the laser frequency from the atomic-transition frequency or the trap radio-frequency confining voltage, they observe hysteresis in the melting and solidification. They model the transition by three-dimensional molecular-dynamic calculations which replicate many of the principal features of the experiments. In particular, the main cause of heating of the ions is a deterministic but chaotic mechanism characteristic of non-linear driven systems. This implies that deterministic chaos rather than stochastic spontaneous emission of photons must be considered instrumental in the crystallization process.

The study of simple phase-transition dynamics in clouds of trapped ions will allow non-equilibrium statistical mechanics to be investigated directly with relatively few particles. Indeed, the changing nature of the transition as the number of particles increases offers new insights for correlated statistical systems. Unlike conventional microscopic solids, systems of laser-cooled ions can be imaged directly and photographed. The fascinating ordered geometrical structures of these fragile, extended crystals reveals the delicate interplay between Coulomb repulsive forces, trap confinement and radiative dynamics. The normal-mode collective vibrations and rotations of the collections of ions are easily visible.

As a bizarre demonstration of the strength of the forces holding the crystal together, off-axis laser cooling can spin two ions around to generate a 'water-mill' pattern in the image fluorescence (see figure). The laser crystals can be further distorted spatially by stray fields indicating how easy they are to disrupt.

The most exciting aspect of the experiment is that the melting and crystallization processes can be studied dynamically. The unexpected hysteresis and bistability observed and the role of deterministic chaos in heating the plasma suggest that this simple laser crystal can shed new light on the importance of nonlinearities in non-equilibrium phase transitions. □

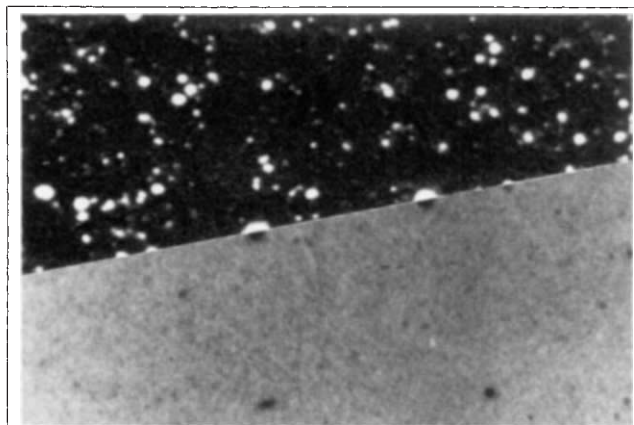
Peter Knight and Richard Thompson are at the Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK.

Astrophysics

Dark matter distortions

J.A. Tyson

THE highly luminous object H1413+117 is a fourfold image, shaped like a cloverleaf, according to Magain *et al.* on page 325 of this issue¹. This is the result of lensing by a massive galaxy bending the light from a more distant quasar. This gravitational lens is one of only a few so far discovered. But importantly, the components in the image are only slightly separated: astronomers have been puzzled by the apparent lack of such tightly focused images.



More lensed quasars are known than lensed galaxies because their great brightness means that they can be seen to greater distances. But new charge-coupled device (CCD) detectors can record images too faint for photographic plates. The figure (after ref. 13) compares a region at the Southern Galactic Pole as recorded by sensitized photographic film (below) and CCD (above). The 20-fold increase in sensitivity reveals that the sky is covered by detectable galaxies at all distances, some of which are probably lensed.

Our view of the Universe is biased in favour of luminous objects. Although dark matter could comprise 90–99 per cent of the Universe by mass, it is by its very nature difficult to study directly. Gravitational lenses provide a vital tool for the study of clumped mass. Einstein showed that the Sun's gravity bends light rays from distant stars, and speculated that a star fortuitously aligned with a distant object would create a ring image². Zwicky argued³ that galaxies would act as gravitational lenses, distorting and amplifying background objects.

Not until 1979 was a gravitational lens discovered — a distant quasar split into at least two images by the gravity of a foreground galaxy⁴. The four detected images that Magain *et al.* report in this issue¹, are separated by only 1 arcsecond and their spectra are consistent with having a single quasar source. The lensing galaxy is too faint compared with the lensed quasar to be detected.

After a decade of searching 4,000 quasars for multiple images, only a few gravitationally lensed quasars have been identified. But in the past year, several

cases of galaxies in the background lensed by foreground galaxies have been discovered. Out of 18 gravitational-lens candidates reported, six are considered to be established and three are very likely. The remaining nine have insufficient or conflicting data.

To confirm the existence of a lens, it is necessary to compare spectra of the images with each other and with the lensing object (if seen). The number of lensed

quasars is expected to increase as the angle between the images decreases⁵, although this has not previously been observed for angles less than 2 arcseconds. The compact multiple image of H1413+117 and the strikingly similar candidate QSO2237+0305 discovered a few years ago^{6–8} reverse this trend. So does the small Einstein ring recently shown on the cover of *Nature*⁹, formed by the exact alignment of source and lens.

The statistics of the lenses are of greatest interest. Various forms of clumped dark matter — from the massive dark

halos associated with galaxy clusters to the small dark objects which could lurk in galaxy halos — should cause testable statistical correlations in the images of the background Universe. The statistical evidence from lensed quasars leads to an interesting conclusion: there is insufficient compact dark matter — black holes, for example — to 'close' the Universe, that is, asymptotically to stop its expansion. Multiple-image lensed quasars would be more frequent were the Universe closed by compact dark matter.

Most of the known lensed quasars are caused by some foreground galaxy dominated by its dark matter, the number being consistent with galaxy masses of about 10^{12} solar masses if the observed number density of foreground galaxies and the number of background quasars are correct. Galaxy lenses offer an independent measure of the galaxy mass distribution. Microlensing — chance lensing by small objects in the halo of the lens galaxy — should also be observable. With luck and much work, detailed studies of lenses could eventually yield new constraints on the Hubble constant, which

measures the present rate of expansion of the Universe.

The chance of a quasar and a foreground galaxy being aligned within the gravitational-lens critical angle (about 1 arcsecond for a galaxy like ours) is reasonably good: out to a redshift of 1 there are over 50,000 galaxies per square degree (redshift, the functional increase in wavelength of emitted light, is used as a measure of astronomical distance). The rarity of strongly lensed background quasars, with redshifts typically of 2, is due mostly to the scarcity of quasars compared with galaxies.

But lensed quasars could be just the brightest examples of gravitational lensing to be found. At the limit of detection, the sky is covered with a veil of nearly overlapping faint galaxies: there are over 150,000 galaxies per square degree that are brighter than 27 magnitude (see figure). Compared with quasar lenses, the probability of galaxy-galaxy lensing is very high. Recently, the stretched images of background galaxies have been found encircling rich galaxy clusters in the foreground^{10,11}, confirming the enormous dark

matter content of galaxy clusters.

The nature of dark matter will be revealed in the way it clumps and prospects are good for a direct measure of the dark-matter distribution. Research in galaxy-galaxy lensing is yielding constraints on the size of galactic dark-matter halos¹² and searches for totally dark gravitational lenses may show how and where dark matter clumps. □

1. Magain, P. *et al.* *Nature* **334**, 325–327 (1988).
2. Einstein, A. *Science* **84**, 506 (1936).
3. Zwicky, F. *Phys. Rev. Lett.* **51**, 290 (1937).
4. Walsh, D., Carswell, R.F. & Weymann, R.J. *Nature* **279**, 381–384 (1979).
5. Turner, E.L., Ostriker, J.P. & Gott, J.R. III *Astrophys. J.* **284**, 1–22 (1984).
6. Huchra, J. *et al.* *Astron. J.* **90**, 691–696 (1985).
7. Yee, H.K.C. *Astron. J.* **95**, 1331–1339 (1988).
8. Schneider, D.P. *et al.* *Astron. J.* **95**, 1619–1628 (1988).
9. Hewitt, J.N. *et al.* *Nature* **333**, 537–540 (1988).
10. Soucail, G., Fort, B., Mellier, Y. & Picat, J.P. *Astr. Astrophys.* **172**, L14–L16 (1987).
11. Lynds, R. & Petrosian, V. *Bull. Amer. astron. Soc.* **20**, 644 (1988).
12. Tyson, J.A. in *Theory and Observational Limits in Cosmology* (ed. Stoeger, W.R.) 441–461 (Specola Vaticana, 1987).
13. Tyson, J.A. *Astron. J.* **96**, 1–23 (1988).

J.A. Tyson is in the Physical Research Laboratory, AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA.

Ecology

The dam busters

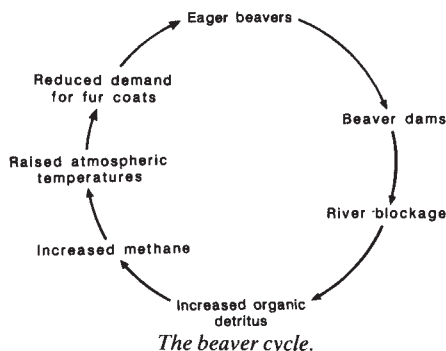
Peter D. Moore

THE recent conclusion by Blake and Rowland¹ that the concentration of methane in the troposphere has risen by 11 per cent in the past 9 years, and the proposal of Pearman and Fraser² that levels have doubled in the past century, are timely reminders of the importance of this gas in the atmosphere. Its potential contribution to the greenhouse effect is currently a cause for concern.

How much of this methane production can be blamed on human activities (through livestock, rice paddies and fossil carbon combustion) is still in dispute and is likely to remain so until more is known about natural sources. Among these, wetlands seem to be of particular importance³. Anoxic conditions are necessary for methane production, so streams and fast-moving rivers are not rich sources, but if a stream becomes blocked and organic sediments accumulate in the resulting body of water, methane can be produced. This consideration has led Ford and Naiman⁴ to examine the role of the beaver (*Castor canadensis*) in methane generation within the boreal forests of Canada, as beavers are known to have a considerable influence on stream hydrology.

Naiman and Melillo⁵ have already demonstrated the extent of the impact of beavers on Canadian forests and their streams, suggesting that between 20 and 40 per cent of second- to fifth-order

streams are affected by beaver activity. They selected sites along tributaries of the St Lawrence River in Quebec for detailed study, some situated at points blocked by beaver dams and others in faster-moving stream sections. Floating gas samplers (inverted glass carboys) were set up at each of 10 sampling stations and aliquots were removed weekly. Methane concen-



tration was determined using gas chromatography. It was impossible to continue measurements following ice establishment, but rates of methane production dropped to 20 per cent of their former values when ice first formed, so this figure was used in order to estimate winter methane flux.

It turns out that the rates of methane production vary considerably from site to site, depending on such factors as temper-

ature and water depth. Values from beaver ponds were up to 33 times higher than those from neighbouring streams with faster flow. Add to this the fact that the area of water surface is on average seven times greater per unit stream length when dammed by beavers, and that up to 40 per cent of streams may be thus affected, and the influence of the beaver on the biome begins to look serious. Excluding natural lakes and peatlands, Ford and Naiman calculate that the beaver is responsible for 92 per cent of the methane production from the river catchments studied, and is a significant influence on an important biogeochemical cycle. It is tempting to wonder whether the beaver is aware that the greenhouse effect will reduce the global demand for fur coats.

But before beginning on a campaign of beaver dam destruction, it is worth looking to see whether there is a natural biological control agent. A contender for this role, according to the work of Reid *et al.*⁶, is the river otter (*Lutra canadensis*), which can enter beaver ponds under the ice in winter by cutting its way through the dams and leaving leaking channels in its wake. These authors suspected that the otter was behaving in this way because of the sizes of the passages in dams and the toothmarks on exposed sticks. They confirmed the passage of otters through beaver dams by tracking them with implanted radio transmitters. The beavers do not begin to repair their dam structures until the ice melts in the spring. Meanwhile, the water levels of the beaver ponds are lowered.

It seems likely that the activity of beavers in creating ponds is raising the carrying capacity of the habitat for otters by enhancing the supplies of fish in winter. But whether the otter is having a marked influence on methane generation is doubtful. In the first place, the dam-busting takes place in winter when methane production is low, and secondly the work of Ford and Naiman suggests that shallower pools have a higher rate of methane production than deep pools. But compared with the world's rice paddies and cattle rumens, not to mention the recently discussed fossil fuel sources⁷, it is doubtful whether the beaver yet constitutes a threat to our climate. □

1. Blake, D.R. & Rowland, F.S. *Science* **239**, 1129–1131 (1988).
2. Pearman, G.I. & Fraser, P.J. *Nature* **332**, 489–490 (1988).
3. Baker-Blocher, A., Donahue, T.M. & Mancy, K.H. *Tellus* **29**, 245–250 (1977).
4. Ford, T.E. & Naiman, R.J. *Can. J. Zool.* **66**, 529–533 (1988).
5. Naiman, R.J. & Melillo, J.M. *Oecologia* **62**, 150–155 (1984).
6. Reid, D.G., Herrero, S.M. & Code, T.E. *J. Mammal.* **69**, 100–107 (1988).
7. Lowe, D.C., Brenninkmeijer, C.A.M., Manning, M.R., Sparks, R. & Wallace, G. *Nature* **332**, 522–525 (1988).

Peter D. Moore is Reader in Ecology in the Division of Biosphere Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

Vision

The thermal limit to seeing

H. B. Barlow

IN BIOLOGY it is rare for a simple question to have a simple answer, so one might expect "Why can't we see dimmer lights?" to require a long and complicated answer. But the paper by Aho and colleagues¹ on page 348 of this issue goes a long way towards proving that a single physical factor, the thermal isomerization of the photosensitive rhodopsin molecules in the retinal rod photoreceptors, is the dominant factor setting the ultimate limit to visual sensitivity.

It was shown almost 50 years ago that the absorption of less than a dozen quanta is sufficient to cause a dim sensation of light in humans². Soon after this, it was argued³ that the reason why several quanta have to be absorbed before a light stimulus can be detected is the irregular thermal activation of the photosensitive molecules, which mimics their activation by single quanta. As there are 10^6 rhodopsin molecules in each receptor cell, thermal activation should happen frequently enough to cause noise at absolute threshold, even if the half-life of rhodopsin were about 100 years.

Hyperpolarizing 'bumps' exactly like those resulting from absorption of light quanta have indeed been shown to occur spontaneously in photoreceptors⁴; from other evidence, they seem to be caused by thermal isomerization of rhodopsin, and the rate at which they occur is close to that required to limit visual sensitivity in dim light.

The original speculation was based on human psychophysics and only established upper limits to the allowable noise. Aho *et al.*¹ now show that the toad *Bufo bufo* has a behavioural threshold about one eighth that of humans — as low as allowed by the known thermal isomerization rate in the toad's receptors. They also show that retinal ganglion cells, which process the signals from the photoreceptors, start to respond reliably at about the same level. This new work, together with that on the photoreceptors, now establishes a firm link between thermal events at a molecular level and their behavioural consequences.

The thermal isomerization rate has a predictable relationship with temperature, increasing about fourfold for a 10 °C rise. Aho *et al.* test this by measuring the threshold for phototactic jumping in the frog *Rana temporaria* over a range of temperatures from 10–20 °C and observe about a fourfold increase of threshold: the molecular rate of photoisomerizations at threshold is roughly equal to the molecular rate of thermal isomerizations. In

other experiments, the relationship also holds for *Bufo bufo* at 15 °C and humans at 37 °C.

This linear relationship gives pause for thought. Although it certainly indicates that the thermal rate is a limiting factor, there are complications. First, when the frog looks at a target, such as a worm, it is the number of thermal events in the target area over the duration of exposure to the target that sets the detection limit, rather than the molecular rate of isomerizations. Second, a square-root rather than a linear relation would be expected. Many other factors besides the molecular rate influence the actual numbers of events occurring during exposure to the target and it is strange that these factors conspire to produce a near-linear relation both in the across-species comparison and for the effects of temperature in a single species. Perhaps there is still more to understand about the problem, but the regular rise of threshold with temperature, together with the fact that the performance is close to the limit imposed by the known rates, would seem to establish the importance of thermal isomerization.

Why is it possible to point to a single dominant physical factor in this case, when the analogous question about other biological limits leads into a maze of interconnecting factors? In the case of the eye, the first point is that it was known where to look. Albert Rose⁵, while developing television cameras, first drew attention to the low quantum requirements of human vision; he was led to that directly from his experience with the comparatively high quantum requirements of the photocathodes of his day. He did not point to thermal noise as the limit of human vision but that was an obvious step because it is thermal emission that limits the ultimate sensitivity of photocathodes.

The second point is that good optics are required for intrinsic thermal noise to limit visual sensitivity. That means, among other things, well-packed and aligned photoreceptors and the appropriate connections with retinal ganglion cells that make suitable connections in the central nervous system. Thus there is also a maze of interconnecting factors in this case, and for thermal noise to stand out implies that evolution can successfully optimize the other factors but can go no further in improving the thermal stability of rhodopsin.

Nor should it be forgotten that basic physics dictates that the physical size of an eye is also important in limiting its ultimate sensitivity. Other things being equal,



Janvier Beltran

PODICEPS gallardoi, the hooded grebe, was discovered in 1974, and lives exclusively in certain types of lagoons on the Patagonian plateau in Argentina between 42 and 49 degrees south, an extremely hostile habitat for any species of grebe. Several factors have been suggested to influence its population size: strong westerly winds; egg and chick predation by kelp gulls; aggression by red-gartered coots; and food availability during the rearing season. The Fundacion Vida Silvestre Argentina, which is directly involved in the monitoring, study and conservation of the hooded grebe, would welcome proposals from researchers interested in following up the study of the species biology and reproduction, who can find his/her own funds for an extended period of three years starting around September 1989. Interested readers should contact the FVSA, Defensa 245/51, P6 (1065) Capital Federal, Argentina. □

both the number of quanta caught from a reflecting object of particular dimensions and the number of photosensitive molecules exposed to these quanta increase as the square of eye size; in contrast, thermal noise increases only linearly because it is proportional to the square root of the number of photosensitive molecules. Hence large eyes are more sensitive, which explains why nocturnal animals have large eyes and why small animals need eyes that are relatively large for their body size.

It is only the ultimate sensitivity, measured in the absence of all light except that from the target, that has been shown to be limited by thermal noise. But it is

obviously a very great advantage to be able to catch and eat your prey in the dawn and dusk when your competitors, and the prey itself, are still blind; perhaps we can now discern the unalterable physical factors that limit the evolution of better visual performance under these conditions. □

1. Aho, A.-C. *et al.* *Nature* **334**, 348–350 (1988).
2. Hecht, S., Schlaer, S. & Pirenne, M.H. *J. gen. Physiol.* **25**, 819–840 (1942).
3. Barlow, H. B. *J. opt. Soc. Amer.* **46**, 634–639 (1956).
4. Baylor, D. A., Matthews, G. & Yau, K.-W. *J. Physiol.* **309**, 591–621 (1980).
5. Rose, A. *Proc. Instn Radio Engrs* **30**, 293–300 (1942).

H. B. Barlow is in the Physiological Laboratory, University of Cambridge, Downing St, Cambridge CB2 3EG, UK.

Ozone depletion

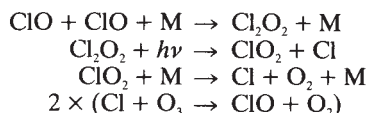
Reactions on ice crystals

John Pyle

RESULTS from several Antarctic observations, including the US Airborne Antarctic Ozone expedition last August to September were discussed at a recent meeting.* They show that the springtime ozone level in the polar stratosphere fell to 40 per cent of the values measured in the 1970s, the strongest depletion yet recorded. Other data confirm the leading role played by heterogeneous chemistry on ice particles in the stratospheric clouds in producing the ozone hole. Measurements from the Arctic stratosphere show ozone depletion could happen there, although no strong effect has yet been observed.

The rapid loss of ozone above the South Pole, first reported in 1985, occurs during about 6 weeks after the return of sunlight in the Antarctic spring but is reversed by December. The depletion has been getting steadily worse each year. The Antarctic has a unique meteorological environment: strong zonal winds produce a cold air mass in the lower stratosphere, chemically isolated from lower latitudes by poor north–south mixing. Ozone depletion is caused in this vortex by a catalytic cycle based on chlorine atoms (see the News and Views article by Brian Thrusch in *Nature* **332**, 784; 1988) originating from man-made chlorofluorocarbons. Ice crystals in the stratospheric clouds that can form in the cold vortex are thought to be sites of crucial heterogeneous reactions that free active chlorine.

The principal catalytic cycle, involving the chlorine monoxide dimer, that destroys ozone is:



where M is any third molecule and $h\nu$ is a photon of light. Measurements of ClO were made last year by an *in situ* resonance fluorescence technique from an ER-2 aeroplane, flying at about 18 km in the lower stratosphere (J. Anderson, Harvard University). Typical flights were from Punta Arenas (58° S) to about 72° S. The most striking feature seen in all the flights was the very high level of ClO concentration, up to 1 p.p.b.v. (part per billion by volume), observed at southernmost latitudes. At the most northern latitudes, the ClO level was approximately constant at 0.1 p.p.b.v.. These values are higher than predicted by numerical models. The sharp gradient in ClO separating the two regimes was found across about 1 degree of latitude. Many other measurements showed similar sharp gradients in constituent concentrations, confirming that there is an isolated, distinct air mass (the 'chemically-perturbed region') over Antarctica.

In August the ozone levels measured by the ER-2 were relatively constant. By mid-September the ozone levels in the region of high ClO were reduced dramatically. Anticorrelations with ClO were seen on both large and small scales. Furthermore, Anderson could infer the Cl_2O_2 mixing ratio in his experiment; all the recorded levels agree well with theoretical predictions.

The role of polar stratospheric clouds (PSCs) in ozone depletion was mere hypothesis in 1986 (Solomon, S. *et al.* *Nature* **321**, 755–758; 1986). Now many of their physicochemical properties are known. For example, the reaction probabilities of several important surface reactions have been measured in the laboratory. Both the ER-2 and DC-8 aircraft were used to identify two populations of particles in the PSCs (D. Fahey, National Oceanic and

Atmospheric Administration (NOAA); L. Poole, NASA). Below about 195 K, micrometre-sized particles (type-I PSCs) containing nitric acid and water were found. These could be frozen droplets of nitric acid trihydrate which, as pointed out by several authors, freezes at higher temperatures than water, providing sites for heterogeneous chemistry. It is supposed that reactions on the surface of these small crystals such as



liberate molecular chlorine which is rapidly photolysed to produce chlorine atoms, the nitric acid being absorbed into the droplet.

At temperatures below about 187 K larger water–ice crystals (type-II PSCs) were found. These larger crystals, which also can absorb HNO_3 , could settle out from the lower stratosphere, leaving it denitrified and dehydrated. The denitrification is essential for the ozone depletion, otherwise excess ClONO_2 would act as a reservoir tying up active chlorine.

Measurements have also been made this year in the Arctic stratosphere. Relatively high column densities of OClO, an indicator of unusual chlorine chemistry, were found above Thule (G. Mount, NOAA). Balloon measurements from northern Sweden (U. Schmidt, Julich Nuclear Research Centre) indicate that the levels of CH_4 and N_2O are very low in the Arctic lower stratosphere and similar to levels above Antarctica. There is no good evidence in the north of strong ozone depletion such as that seen in Antarctica but some of the conditions are evidently favourable for depletion to occur.

Measurements from the ER-2 in the north taken outside the vortex confirm that the Arctic may be perturbed. W. Brune (Harvard University) reported that the concentrations of ClO are much higher than predicted by models. These measurements could also have important implications globally: there is more active chlorine in the lower polar stratosphere than models currently predict. How much of this chlorine is then transported towards the Equator in the lower stratosphere and, therefore, how inadequate are present models for global prediction? It seems likely that models are particularly poor in this most important region. Measurements which could support this suggestion were presented by A. Tuck (NOAA), who showed evidence of dehydration of the lower stratosphere at latitudes north of Punta Arenas. If this air had previously been within the vortex, other constituent concentrations too should be quite untypical for mid-latitudes. □

John Pyle is in the Department of Physical Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

* Polar ozone workshop, Aspen, Colorado, 9–13 May 1988.

Astrophysics

Missing link of pulsar evolution

Charles Bailyn

THE discovery¹ of the eclipsing millisecond pulsar PSR1957+20 has sent theoreticians scurrying to find a place for this remarkable object in the complex life-cycle of radio pulsars and their cousins, the low-mass X-ray binaries. E.S. Phinney *et al.*², W. Kluzniak *et al.*³, and E.P.J. van den Heuvel and J. van Paradijs⁴ show that the pulsar represents an evolutionary link, hitherto only hypothesized, between low-mass X-ray binaries and isolated pulsars observed to have millisecond periods.

Pulsars are rotating neutron stars which beam radiation towards an observer once each spin period; most have pulse periods of about 1 second but a few have periods up to a thousand times shorter. These high rotation speeds are thought to result from a period of accretion earlier in the life of the neutron star. In this model, gas from a nearby binary companion star accumulates in an accretion disk around the neutron star. As this material subsequently settles onto the neutron star it brings with it a high specific angular momentum, spinning up the neutron star to the short spin periods required for millisecond pulsations.

The gravitational energy released during the accretion phase manifests itself in high X-ray luminosities and the system is observed as a bright X-ray source, not a radio pulsar. After the termination of accretion, the neutron star is 'uncovered' and emerges as a radio pulsar. How and why accretion ends and, in the case of isolated (non-binary) millisecond pulsars, how the binary companion is removed, have not until now been understood. PSR1957+20 has provided dramatic confirmation of an important process in this regard, being an intermediate evolutionary stage between low-mass X-ray binaries (LMXBs) and isolated millisecond pulsars.

Companion

The size of the eclipsing region of the companion of PSR1957+20 is estimated to be 0.7 times that of the Sun. Changes in the pulse timing as the pulsar approaches and recedes from us in its binary orbit give a velocity and size for the orbit, and thus an estimate of the companion's mass of 0.02 solar masses. These measurements of the size and mass of the companion lead to an apparent paradox, because an object of these dimensions would be ripped apart by the tidal field of the pulsar itself.

The resolution to this problem is contained in two papers^{5,6} by Kluzniak's colleagues, submitted to the *Astrophysical*

Journal a few months before the discovery of PSR1957+20. They suggest that energy released by the gradual spin-down of a millisecond pulsar is transferred to a binary companion in the form of megaelectron-volt photons, which give rise to a strong wind from the companion. Such a wind would suffice to block the radio emission from the pulsar to a distance many times the radius of the companion. In this case the companion star itself could be compact enough not to be tidally disrupted. Thus PSR1957+20 belongs to that select group of astronomical objects for which theoretical explanations existed before their discovery.

The papers by Phinney *et al.*², Kluzniak *et al.*³ and van den Heuvel and van Paradijs⁴ apply this model directly to PSR1957+20 and explore various evolutionary consequences. The authors agree that the pulsar luminosity suffices to drive the wind required to create the observed eclipses, and also that within 10⁶ years, the companion will be completely evaporated. In this time, the pulsar spin-down will not be significant, so PSR1957+20 will evolve into an isolated millisecond pulsar.

Having concurred on the present and future status of the eclipsing pulsar, Phinney *et al.*² and Kluzniak *et al.*³ differ over the past evolution of the system. Phinney *et al.*² suggest that the neutron star's companion (the secondary) in the precursor LMXB system was a main-sequence star of 0.6 solar masses and that the system had a 5-hour orbital period. If mass transfer suddenly stopped at this point, perhaps because of changes in the internal structure of the main-sequence star that occur at this mass, the uncovered pulsar would begin to evaporate the secondary. The mass and angular momentum driven off from the secondary would then result in a system with the parameters of PSR1957+20.

In contrast, Kluzniak *et al.*³ begin with a secondary of less than 0.1 solar masses and an orbital period of about 1 hour. As the secondary becomes less massive, gravitational radiation becomes less effective at driving mass transfer. Traditionally, the resulting gradual decrease in luminosity has been thought to mark the demise of LMXBs as observable objects. But Kluzniak *et al.* suggest that mass transfer may be driven by megaelectron-volt γ -rays which are produced at the boundary between the accretion disk and the magnetosphere of the neutron star. This radiation creates a wind from the secondary, which is accreted into a disk and

from there onto the neutron star without loss of angular momentum, driving the system apart as the mass ratio becomes more extreme. Eventually this source of accretion also becomes small and the pulsar becomes uncovered in a system close to the present parameters of PSR1957+20. According to this scheme all dying LMXBs will be reborn as millisecond pulsars.

Cataclysmic variables

van den Heuvel and van Paradijs⁴ apply the wind-evaporation idea to another question in LMXB evolution. They note that cataclysmic variables, which are similar to LMXBs except that they have white dwarf primaries instead of neutron stars, are known to have a range of periods (2–3 hours) in which mass transfer apparently ceases. This is probably also true for LMXBs: none with a period in this range is known. When the mass transfer turns off in an LMXB, the uncovered pulsar quickly evaporates the companion. Thus, although cataclysmic variables with orbital periods below 2 hours are thought to have evolved through the gap by gravitational radiation, this process cannot occur in LMXBs; instead these become millisecond pulsars. There is some observational support for this, in that the few LMXBs with periods below the gap are thought to have degenerate white-dwarf secondaries and different evolutionary paths altogether.

These three papers^{2–4} make clear that the evolution of LMXBs and millisecond pulsars must now be considered as one subject, involving the same physical processes. The vast energy output of LMXBs means that we have identified all of these objects contained in our Galaxy. Equally certainly, we are only seeing the tip of a vast iceberg of millisecond pulsars, as the discovery of five millisecond pulsars in globular clusters over the past year confirms.

Thus future progress in this area could be dominated by studies of the millisecond pulsars, both in and out of binary systems. We can also expect increasing understanding of PSR1957+20 itself — recently an optical counterpart has been discovered⁷ and the pulse period derivative has been determined (Fruchter, A.S., personal communication). □

1. Fruchter, A.S., Stinebring, D.R. & Taylor, J.H. *Nature* **333**, 237–239 (1988).
2. Phinney, E.S., Evans, C.R., Blandford, R.D. & Kulkarni, S.R. *Nature* **333**, 832–834 (1988).
3. Kluzniak, W., Ruderman, M., Shaham, J. & Tavani, M. *Nature* **334**, 225–227 (1988).
4. van den Heuvel, E.P.J. & van Paradijs, J. *Nature* **334**, 227–228 (1988).
5. Ruderman, M., Shaham, J. & Tavani, M. *Astrophys. J.* (in the press).
6. Ruderman, M., Shaham, J., Tavani, M. & Eichler, D., *Astrophys. J.* (in the press).
7. Kulkarni, S.R., Djorgovski, S. & Fruchter, A.S. *Nature* (in the press).

Charles Bailyn is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

Excitation-contraction coupling

Proteins that bridge the gap

William S. Agnew

WHAT are the molecular steps by which electrical impulses induced on the surface of a muscle cell are converted to precisely controlled physical movements of contracting muscle fibres? This question has long drawn attention to the mechanisms at work in a narrow space between an intracellular extension of the sarcolemma, the t-tubule, and the flanking terminal cisternae of the intracellular membrane system, the sarcoplasmic reticulum (SR)¹.

Postsynaptic depolarizations are normally amplified and propagated along the muscle cell surface as Na⁺-dependent action potentials, travelling to the cell interior via the t-tubule system. These depolarizations are sensed by elements in the t-tubule membrane and transmitted to the closely apposed SR to initiate Ca²⁺ release and contraction. The gap between these membranes is about 160 Å and is bridged at regular intervals by large particles generally called the SR 'feet'². Evidence is rapidly accumulating that the mechanism whereby t-tubule depolarization commands SR Ca²⁺ release resides at these sites and involves structurally elaborate proteins. These include the dihydropyridine (DHP) receptors in the t-tubule, which may be the voltage-sensors³, and ryanodine receptors, which seem to act as the SR Ca²⁺ release sites.

DHP receptors have been implicated both as Ca²⁺ channels and as voltage sensors for excitation-contraction coupling. Compounds such as nitrendipine and congeners of verapamil are potent antagonists of L-type Ca²⁺ channels. These compounds block slow Ca²⁺ currents in muscle and prevent contraction^{3,4}. A tempting, if simplistic, explanation is that Ca²⁺ influx might act as an internal messenger to activate SR release, with the channel-gating mechanism serving as the voltage sensor. But this hypothesis is clearly inadequate, despite growing evidence that the DHP receptor may be the voltage sensor, and that it exhibits many properties which might be expected for a Ca²⁺-channel protein. The reason is that Ca²⁺-channel function *per se* seems not to be required for coupling^{3,4}. Ca²⁺ currents are slower and longer lasting than activation of contraction. Their elimination by selective removal of extracellular Ca²⁺, or blockade with Cd²⁺, does not prevent coupling. Indeed, recent experiments demonstrate that coupling occurs even when Ca²⁺ currents are driven outwards⁵. Furthermore, the number of Ca²⁺ channels required to account for the currents is more than 20-fold lower than

the number of DHP receptors.

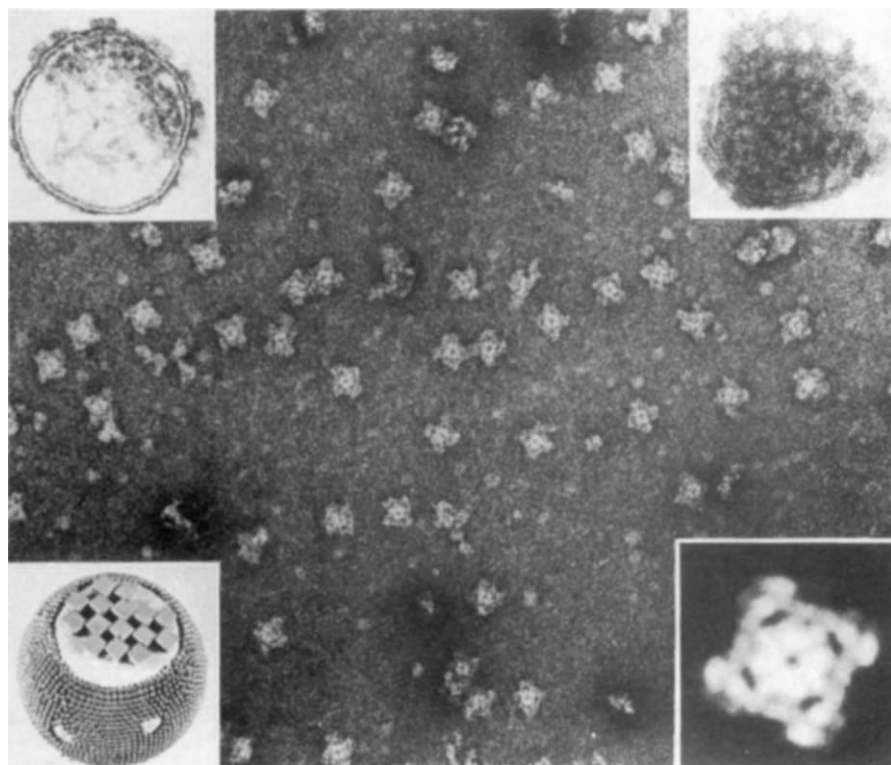
Pursuing the voltage sensor in a series of meticulous experiments, Rios and co-workers have studied the effects of dihydropyridines on t-tubule membrane charge movement closely associated with Ca²⁺ release^{4,6}. These charges can move according to one of two modes: in a state designated 'charge one', the voltage-dependent movement from resting to active configurations correlates linearly with SR Ca²⁺ flux into the myoplasm. When shifted into the 'charge two' state, movement is no longer correlated with release. Dihydropyridines bind preferentially to the ineffective charge two state, simultaneously reducing the amount of charge one and the Ca²⁺-release rates.

A new twist in the story concerns the effects of extracellular ions^{5,6}. Although eliminating Ca²⁺ alone does not prevent coupling, removal of all metal ions shifts the charge into the inactivated charge two state, again proportionately reducing SR Ca²⁺ flux. Rios and co-workers have discovered an intriguing selectivity sequence among both monovalent and divalent ions

for maintaining the charge one state.

Models of ionic permeation through L-type channels (see ref. 7) suggest that there is binding to two sites, separated perhaps by 11 Å. Ions which bind weakly are predicted to produce large currents, whereas those binding tightly are poorly conductive and may even block the channel. The sequence discovered by Rios and co-workers for sustaining charge one corresponds almost precisely to that for the binding of monovalent and divalent ions rather than their ability to carry current through Ca²⁺ channels. Thus, even if the receptor does not form an obligatory conductive channel², it exhibits a binding site remarkably similar to that in the pathway of L-type channels, which must be occupied for coupling.

The molecular picture of the DHP receptor which is emerging from several laboratories is of a large complex comprising four peptide subunits (see, for example, ref. 8). The receptor peptide has a relative molecular mass (*M_r*) of 212,000 (212K), estimated from the cDNA sequence⁹. It binds dihydropyridines, is phosphorylated by ATP and is very similar to the large peptides of voltage-sensitive Na⁺ channels. It seems very likely this molecule is voltage-sensing; it may be a Ca²⁺ channel or simply have descended from one. It may carry out both roles, depending on its conformation or



Images of the ryanodine receptor. Negative stained particles such as those in the central field were used to generate the enhanced image in the lower right-hand insert. Available evidence suggests this ~300-Å diameter particle derives from four copies of a 400–450K peptide. The upper right insert shows a tangential view of regular arrays of these particles in a cisternal membrane; and the thin section at upper left illustrates how the particles extend ~120 Å from the SR bilayer towards the t-tubule. The model at lower left portrays the terminal cisterna, with organized array of ryanodine receptor and tight arrays of Ca/Mg ATPase. (From ref. 14, courtesy of S. Fleischer.)

state of phosphorylation^{2,9}.

Another, apparently larger glycopeptide and two smaller peptides ($M_s=30K$ and $50K$) are non-covalently associated with the DHP-binding subunit (see, for example, ref. 8). The glycopeptide is being cloned, and the amino-acid sequence which can be deduced from the cDNA sequence should assist in predicting its arrangement in the molecular assembly. The remarkable resemblance of the DHP-receptor peptide to Na^+ -channel peptides suggests that it folds into four membrane-spanning domains arranged in a pseudo-tetramer in the bilayer. How could another subunit of similar size be accommodated? It could form a companion structure, perhaps in a dissimilar conformation. A loosely paired organization would explain the small size of the irradiation inactivation target for the DHP receptor ($M_r=210K$)¹⁰. This notion seems a little less *ad hoc* in the light of dramatic new information about the ryanodine receptor with which it may interact.

Ryanodine is a neutral plant alkaloid which causes the dumping of SR Ca^{2+} stores. When heavy SR membranes are fused into planar bilayers, a multi-state cationic channel is observed which is highly conductive to Ca^{2+} , is activated by micromolar Ca^{2+} and by millimolar ATP, and blocked by calmodulin and ruthenium red. This channel is locked into a conducting substate by ryanodine (see ref. 11). Thus the ryanodine receptor, exclusively associated with the terminal cisternae, seems to be the outstanding candidate for the Ca^{2+} -release channel.

Biochemical isolation of the detergent-solubilized ryanodine receptor reveals a very large particle of M_r about 1,800K (refs 12, 13). The protein is apparently formed of four identical peptides of 400K–450K. When reconstituted into artificial bilayers, the full range of conductance and pharmacological properties attributed to the ryanodine receptors in heavy SR can be reproduced.

Remarkably, when this receptor was examined under the electron microscope, investigators found themselves staring at the SR foot^{12–14}. Fleischer and co-workers have recently performed a detailed examination of the morphology of the particle¹⁴. As illustrated in the figure, it displays an elaborate 4-fold symmetry, and is inserted at one end into the cisternal membrane. Extending more than 120 Å into the triadic gap is a large clover-leaf or quatrefoil structure, which seems very likely to associate with the t-tubule via the DHP-receptor complex.

Among the implications of these findings, assignment of the Ca^{2+} release site to the SR feet, the one structure which physically spans the gap, reinvigorates the hypothesis of Chandler and co-workers¹⁵ that direct conformational coupling links the voltage-sensor in the t-tubule to the

SR channel¹⁵. (Reconstituting such a postulated intermembrane coupling could be feasible, perhaps by driving the opening of ryanodine receptors with Ca^{2+} channel agonists in preparations where intermolecular associations are retained). The extended quatrefoil surface which the ryanodine receptor presents to the t-tubule makes the involvement of a 'dimeric' DHP receptor, or a pair of such assemblies, seem more inviting.

Although these discoveries legitimize a discussion of direct allosteric coupling mechanisms, other phenomena must be considered in the overall cycle of events. Demonstration that Ca^{2+} readily activates conductance of the ryanodine receptor has already stimulated a re-examination of the importance of Ca^{2+} -activated Ca^{2+} release, together with the question of how this conductance is terminated to permit Ca^{2+} reuptake⁶.

Furthermore, potential roles for the inositol triphosphate cascade continue to present themselves. In SR membranes from frog skeletal muscle, as well as smooth muscle, inositol 1,4,5-trisphosphate has now been shown to activate ryanodine-sensitive Ca^{2+} conductances¹⁶. Does this relate directly to coupling or to modulation of the effects of conformational activation by the DHP receptor? Given the importance of maintaining control of Ca^{2+} levels at all stages of the excitation-contraction coupling, could this cascade be involved in maintaining Ca^{2+} homeostasis, including recruitment of extracellular Ca^{2+} into the intracellular cycle? If this turns out to be true, is there a role for other products of the cascade, and how are the responsible enzymes activated? Needless to say, the fascinating paradoxes surrounding excitation-contraction coupling remain unresolved, but we can assert that our state of confusion has advanced to a new level of sophistication. □

1. Somlyo, A. *Nature* **316**, 298–299 (1985).
2. Franzini-Armstrong, C.M. & Nunzi, G. *J. muscle Res. Cell Motil.* **4**, 233–252 (1983).
3. Agnew, W.S. *Nature* **328**, 297 (1987).
4. Rios, E. & Brum, G. *Nature* **325**, 717–720 (1987).
5. Brum, G., Fittz, R., Pizarro, G. & Rios, E. *J. Physiol., Lond.* **398**, 475–505 (1988).
6. Pizarro, G., Fittz, R. & Rios, E. *News Physiol. Sci.* (in the press).
7. Tsien, R.W., Hess, P., McClesky, E.W. & Rosenberg, R.L. *A. Rev. Biophys. biophys. Chem.* **16**, 265–290 (1987).
8. Leung, A.T. *et al.* *J. biol. Chem.* **263**, 994–1001 (1987).
9. Tanabe, T. *et al.* *Nature* **328**, 313–318 (1987).
10. Norman, R.I., Borsotto, M., Fosset, M., Lazdunski, M. & Ellory, J.C. *Biochem. biophys. Res. Commun.* **111**, 878–883 (1983).
11. Imagawa, T., Smith, J.S., Coronado, R. & Campbell, K. *J. biol. Chem.* **262**, 16636–16642 (1987).
12. Lai, F.A., Erickson, H.P., Rousseau, E., Liu, Q.-Y. & Meissner, G. *Nature* **331**, 315–319 (1988).
13. Inui, M., Saito, A. & Fleischer, S. *J. biol. Chem.* **262**, 1740–1747 (1987).
14. Saito, A., Inui, M., Radermacher, M., Frank, I. & Fleischer, S. *J. Cell Biol.* (in the press).
15. Chandler, W.K., Rakowski, R.F. & Schneider, M.F. *J. Physiol., Lond.* **254**, 285–316 (1976).
16. Suarez-Isla, B.A. *et al.* *Biophys. J.* (in the press).

William S. Agnew is in the Department of Cellular and Molecular Physiology at Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510-8026, USA.

Daedalus

Counting stars

STELLAR distances are estimated in many ways. The most basic method, which calibrates all the others, is that of orbital parallax: measuring the tiny shift in a star's position as seen from opposite sides of the Earth's yearly orbit round the Sun. Sadly, this method will only work with rather close stars — within a couple of hundred light years or so. Daedalus now has a new method.

At some point in its orbit, he says, the Earth is closest to the target star; six months later it is furthest away. So the star should fluctuate in apparent brightness on a yearly cycle, with a maximum when the Earth is closest, and a minimum when it is most distant. The effect would be very small. But Daedalus points out that the intensity of light can nowadays be measured with absolute precision, by photon counting. This is already used in many telescopes, and could easily be adapted to integrate the light from a single star.

The snag, of course, is that light is inherently 'noisy'. A photon count is reproducible only to within about the square root of the total count. Even so, Daedalus calculates that a telescope of ten square metres cross-section, counting for a fortnight, could measure the brightness of a star like the Sun sufficiently precisely to register its yearly brightness cycle from 500 light years away. Bigger mirrors, longer integration times, and brighter target stars could extend the technique to far greater distances still.

No earthbound instrument, peering through the turbulent soup of the atmosphere, could do this job. A space-borne telescope is essential, but need not be very expensive. A photon-counting detector does not need a sharp stellar image, so a very crude mirror should suffice: just good enough to separate the target star from its neighbours. Daedalus envisages a plastic-film balloon, inflated in the zero-gravity vacuum of space into just the right curvature for its aluminized back face to form a simple concave mirror. It will be roughly attitude-stabilized to keep its imperfect image of the chosen star on the photomultiplier array at the instrument's focus. The count would be relayed continuously to Earth. The whole thing should not tax the state of the art in any direction, and might be an ideal European contribution to space technology.

Daedalus would be particularly pleased to see his new instrument used to measure the distances of quasars. These are now so universally assumed to be immensely bright and immensely distant that the proof of the converse belief would set a corrective cat among a flock of rather self-satisfied pigeons.

David Jones

Improved stability for biological nomenclature

SIR—The pressure from the users of names of organisms on taxonomists to produce more stable systems of names is increasing¹⁻⁴. Names change for one of two reasons: the strict application of the rule of priority or other nomenclatural caveats of the appropriate International Code; or new knowledge on the circumscription, rank or position of a taxon. Changes of the latter kind result from the advancement of scientific knowledge and are of value because they illuminate not only taxonomic relationships but other features, such as physiological and biochemical attributes. Changes for nomenclatural reasons alone, in contrast, benefit no-one.

In perpetuating the present system, taxonomists are failing to satisfy their consumers⁵ and it is therefore not surprising that support for taxonomic research and services is limited. The instability of names is a severe handicap for such consumers as developers of data retrieval systems, and for health, trade, conservation and quarantine authorities in drawing up legislation, regulations and property rights protection.

Bacteriologists overcame this problem in 1980 by the adoption of a new starting date for nomenclature and the publication of an 'Approved List' of names^{6,7} which reduced the number of species names from about 30,000 to only 2,500. In the case of groups covered by the International Code of Botanical Nomenclature (ICBN), about 36,500 generic and 400,000 species names are in use, out of about 79,000 generic and 1,700,000 species names published. It has been suggested that approved lists of names are issued at five-yearly intervals⁸; that a list of currently accepted names of the world's flora be produced which could be accorded some specially protected nomenclatural status⁹; and in zoology that names in particular books or papers be granted a protected status^{10,11}.

Proposals to introduce formal procedures for the registration of newly published names¹² were debated during the XIV International Botanical Congress in Berlin in July 1987. A special committee on registration was established which is to report to the next Congress in Tokyo in 1993. Such a process would not, however, overcome the instability caused by the

repeated re-introduction of long-forgotten names. The International Union of Biological Sciences (IUBS), with the support of the International Association for Plant Taxonomy (IAPT), sponsored an international meeting at Kew on 22–23 April 1988 to consider the feasibility of the production of lists of names in current use for all groups of organisms covered by the ICBN, namely living and fossil flowering plants, ferns, mosses, hepatics, algae, cyanobacteria, fungi (including lichens) and certain protoctists. The meeting was attended by 23 specialists including key personnel associated with the current cataloguing of names (that is, the *Index Algarum*, *Index of Fungi*, *Index Kewensis*, *Index Muscorum*, *Index Nominum Genericorum*), together with representatives of selected user groups. A full report of this meeting will appear in both *Biology International* and *Taxon*, but its key conclusions are outlined in this letter.

The preparation of lists of names in current use was agreed to be a worthwhile objective in itself, and, if such lists were accorded protected nomenclatural status, this would almost entirely eliminate the majority of name changes due to nomenclatural reasons. This objective is now technically feasible for the ~ 36,500 generic names in use for all groups covered by the ICBN, given the machine-readable and card files which have already been compiled. The starting point for this list will be the IAPT *Index Nominum Genericorum* database held at the Smithsonian Institution, and it should be published in 1991. The situation for the ~400,000 current species names varies markedly and separate lists will have to be prepared for each group such as legumes, mosses and yeasts. Pilot studies are now feasible, given the necessary resources.

It was decided that the IUBS, through its Commission on the Nomenclature of Plants, be asked to establish a special committee on names in current use. This should make detailed proposals to the next International Botanical Congress about granting special status to the lists of generic names and the appropriate mechanisms for updating these lists, as well as about procedures for the preparation and adoption of species names lists. This committee would work in collaboration with the existing committee which is considering the question of the registration of newly published names.

If the necessary funding can be obtained, there is now a plan which, if accepted by the biological community at large, would materially improve the stability of names of all organisms covered by the International Code of Botanical Nomenclature. The participants in the Kew meeting wish to encourage a lively

debate on this matter, and invite comments from both users and taxonomists, which will be considered by the proposed special committee responsible for both the production of the generic and pilot species lists, and for the preparation of detailed proposals for decision at the 1993 International Botanical Congress.

D.L. HAWKSWORTH
CAB International Mycological Institute,
Kew, Surrey TW9 3AF, UK

Seal death

SIR—During the past two months outbreaks of an acute disease have occurred among harbour seal (*Phoca vitulina*) populations in Scandinavian, German and Dutch waters. The most prominent signs of the disease are acute pneumonia in young and adult animals, and abortion in pregnant females. The mortality rate is 5–25%. Cases were first observed in the Danish Kattegat area, where about 200 aborted seal pups and hundreds of dead juvenile and adult animals were found along the coast. The disease gradually spread to the harbour seal populations in the Danish, German and Dutch Wadden Sea.

Epizootiological and post-mortem findings pointed towards an infectious agent as the cause of the outbreaks but the possible role of other factors was discussed during a meeting organized by the Ministry of Environment of the State of Sleeswijk Holstein together with G. Heidemann at the University of Kiel. Special attention was given to the possibility of a triggering role for certain environmental pollutants¹ and of the drastically increased concentrations of certain algae in the same area during the same period. No clear indications for the involvement of these factors were found. Instead there is evidence that one or more viruses are the cause. When virus isolation procedures, using seal primary kidney cell cultures, were applied to the lungs and other organs of 35 dead animals from the Danish, German and Dutch waters, a herpesvirus was isolated from eight animals by P. Häve of the State Veterinary Institute for Virus Research in Lindholm, and ourselves.

Four years ago, we isolated a herpesvirus, *Phocid herpesvirus-1*, during an outbreak of a similar disease with high mortality in baby seals nursed in the Seal Rehabilitation and Research Center in Pieterburen (The Netherlands)². This virus was characterized as a novel member of the Alphaherpesvirinae subfamily and shown to be the causative agent of the disease. Since then, we have shown that neutralizing antibodies to the virus are present in many different pinniped species in different areas, indicating that this, or a closely related virus, is widespread³.

The herpesvirus isolated during the recent outbreak is being compared with

1. Erzincliglu, Y.Z. & Unwih, D.M. *Nature* **320**, 687 (1986).
2. Barnett, J.A. *Nature* **322**, 599 (1986).
3. Rose-Chwatt, L.J. *Nature* **325**, 114 (1987).
4. Rossmore, H.W. *Int. Biodet.* **23**, 325–327 (1988).
5. Hawksworth, D.L. & Bisby, F.A. in *Prospects in Systematics* (ed. Hawksworth, D.L.) 3 (Clarendon, Oxford, 1988).
6. *Nature* **283**, 511 (1980).
7. Sneath, P.H.A. in *Biological Nomenclature Today* (ed. Ride, W.D.L. & Younes, T.) 36–48 (IRL, Oxford, 1986).
8. Barnett, J.A. *Nature* **325**, 384 (1987).
9. Brummitt, R.K. *Taxon* **36**, 78–81 (1987).
10. Cornelius, P.F.S. *Bull. zool. Nom.* **44**, 79–85 (1987).
11. Manuel, R.L. *Bull. zool. Nom.* **45**, 144–145 (1988).
12. Greuter, W. *Taxon* **35**, 816–819 (1986).

Phocid herpesvirus-1, but we are also studying a second virus that is cytopathic for seal kidney cells and cells from a number of other species and which was isolated from the lungs of 20 out of 22 dead seals investigated. By negative contrast electron microscopy this virus has been tentatively classified as a member of the Picornaviridae family.

In the past three years we have applied virus isolation procedures to lungs of 20 normal seals and 15 with signs of acute pneumonia. No virus has been isolated from the former animals but 11 of the animals with pneumonia yielded *Phocid herpesvirus-1*. These data indicate that the newly isolated picorna-like virus, or *Phocid herpesvirus-1*, or both viruses, have probably caused the outbreaks of disease among the seal populations. Serological investigations on convalescent seals will be conducted in an attempt to confirm this hypothesis.

A. D. M. E. OSTERHAUS
Veterinary Advisory Committee,
Seal Rehabilitation and Research Center,
Pierburen, The Netherlands

1. Reijnders, P. J. H. *Nature* **324**, 456–457 (1986).
2. Osterhaus, A. D. M. E. *et al. Arch. Virol.* **86**, 239–251 (1985).
3. Vedder, L., Zarnke, R., Spijkers, I. & Osterhaus, A. *Abst. Seventh Biennial Conf. Biology Marine Mammals* 5–9 December 1987, Miami, Florida.

Effect of bedslope on desert sand transport

SIR—Hardisty and Whitehouse¹ report most interesting field data concerning the wind transport of sand when this is either assisted or retarded by bedslope. We believe that these data will have considerable value in developing and testing adequate theories of aeolian transport. However, their discussion appears to take little account of some features of the process which are reasonably well established. In consequence, they attribute the effect of bedslope to a “new sand transport process”, whereas it can be attributed more helpfully to the influence on known processes of a gravitational field which is not normal to the bed.

Grains are indeed dislodged predominantly by the collision of saltating grains. Much is now known about the outcome of

collisions^{2,3}, which seem to give rise to two distinct populations of launched grains. Ricochet of the incident grain usually occurs and is vigorous enough to enter another saltation; meanwhile a small number of grains (up to twelve or so) are dislodged and undertake much smaller trajectories which Haff has called “reptations” (see ref. 5). The angle of launch has a considerable range (Fig. 1), a significant proportion of dislodged grains emerging with an upwind velocity component. The saltating grains are essentially wind-driven and follow paths which are very similar for sloping and horizontal beds. Reptating grains, however, travel at levels at which grain-laden boundary-layer flow is very sluggish, and are controlled predominantly by gravity. They account for a substantial proportion of the transport rate. It is to be expected, therefore, that a change, relative to the bed, of the direction of the gravitational field will affect the transport rate much more than is predicted by the quoted equation (5), in which internal shear stress of the grain mass is always resisting gravity.

The effect of bedslope on transport rate can be understood only in terms of the detailed consequences of inter-saltation collision. The data obtained by the authors promise to assist the development of such understanding by extending the circumstances in which grain transport models can be tested. Models already exist (see, for example, ref. 5) which could with little adaptation be used to predict changes of transport rate with θ , such as those reported.

B.B. WILLETTTS
M.A. RICE

Department of Engineering,
Kings College,
University of Aberdeen,
Aberdeen AB9 2UE, UK

HARDISTY AND WHITEHOUSE REPLY—Willetts and Rice raise three important points about our paper¹ on desert sand transport to which we would wish to respond. These points concern the sand transport data, the semantics of the subject and finally the source of the energy which drives the impact-induced gravity flow (IIGF) which we postulated to explain the difference between our results and traditional theory.

Dealing firstly with the data, we agree that our results should be used by ourselves and others to test new theoretical developments. To this end we have transferred all our measurements in ASCII format to PC disk. Copies of the disk will be available from the address below.

Semantically we are also in agreement with the papers referenced by Willetts and Rice, which use the word saltation for those grains moving on a trajectory above the bed which is affected by the wind. This distinguishes such grains from those which

are said to be reptating; this term being used to describe those grains which are moved above but close to the bed by an impactor and are unaffected by the shearing wind. Both saltating and reptating grains will of course return to the bed at a point which is separated from their starting point by a distance which varies with the bed gradient. Such is the result of the geometry of the grain path and the distribution of jump lengths is well evidenced by, amongst others, the work of Willetts and Rice.

These considerations lead to the final and most important point, which concerns the question of whether the IIGF which we postulated is indeed a fundamentally different process or, as Willetts and Rice appear to be contending, simply reptation on a slope. Without repeating the description which we gave in our paper, we are suggesting that we have observed a different process. IIGF is more closely linked with the observations of grains vibrating about their niche positions during subaqueous initiation on a horizontal bed, which were reported by Gessler⁶ and Yalin⁷. These vibrations are produced by fluid shear, whereas we postulate that impacting grains can produce a similar response. Anderson⁵ (page 945) referred to such when he noted that less than 1% of the impactors' kinetic energy is transferred to the ejected, saltation and reptation grains, and that the remaining energy induced ‘local transient dilation of the bed, inelastic deformation of bed grains, and frictional rotation of the bed grains’. We are suggesting that under a well-developed saltation flux, these mechanisms greatly reduce the shear strength of the surface grain layer so that a type of gravity flow can occur on non-horizontal beds. The process is different from reptation because the moving grains do not leave the sand surface and are not impelled by direct impact. It is difficult to envisage how the type of simple changes to existing, kinetic models to which Willetts and Rice refer will predict these effects. Our empirical equation (6) provides a quantitative estimate of the magnitude of the sum of all of these processes, and the next step is a more rigorous theoretical model.

J. HARDISTY
R.J.S. WHITEHOUSE

Department of Geography,
The University of Hull,
Hull HU7 6RX, UK

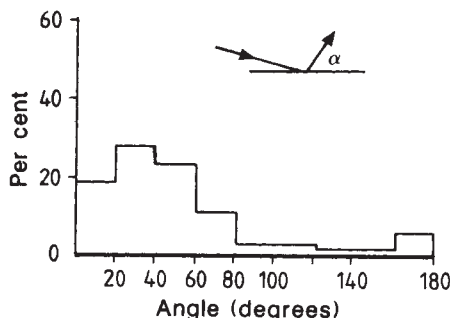


Fig. 1. Distribution, for dislodged grains, of the take-off angle, α , referred to the horizontal bed (see inset).

1. Hardisty, J. & Whitehouse, R.J.S. *Nature* **332**, 532 (1988).
2. Willetts, B.B. & Rice, M.A. *Memoirs No. 8* (Department of Theoretical Statistics, Aarhus University, Denmark, 1985).
3. Willetts, B.B. & Rice, M.A. *Acta mechanica* **63**, 255–265 (1986).
4. Mitha, S., Tran, M.Q., Werner, B.T. & Haff, P.K. *Acta mechanica* **63**, 267–278 (1986).
5. Anderson, R.S. *Sedimentology* **34**, 945–946 (1987).
6. Gessler, J. in *River Mechanics* Ch. 7 (ed. Shen, H.W.) (Fort Collins, 1971).
7. Yalin, M.S. *Mechanics of Sediment Transport* (Pergamon, Oxford, 1972).

Out of this nettle, danger . . .

David L. Sills

Searching for Safety. By Aaron Wildavsky. *Social Philosophy and Policy Center, Bowling Green State University*: 1988. Pp.253. Distributed by Transaction Books, New Brunswick, New Jersey/Ohio Press, Oxford. Hbk \$32.95, £25.30; pbk \$16.95, £11.25.

Searching for Safety is the second book about risk-benefit analysis by Aaron Wildavsky, a distinguished professor of political science at the University of California, Berkeley; the first, *Risk and Culture*, was written with the equally distinguished British anthropologist Mary Douglas. Wildavsky's second excursion onto these difficult seas is laudable, because many of the risks of modern life are associated with technologies that we can neither totally reject nor blindly accept. His results are uneven, but on balance stimulating and useful.

A careful reading of *Searching for Safety* provides a set of questions to ask of the daily onslaught of press and television reports on contemporary medical, environmental and energy debates. Moreover, the book presents a dramatic thesis: "the secret of safety lies in danger" (p. 205). On the same page, Wildavsky approvingly quotes Peter Huber, who quotes Hotspur on safety. Unfortunately, for Wildavsky's case, he does not quote enough of Hotspur's soliloquy — which in fact sets the framework for Wildavsky's thesis:

Hotspur [reading from a letter]: "The purpose you undertake is dangerous" — why that's certain: 'tis dangerous to take a cold, to sleep, to drink; but I tell you my lord fool, out of this nettle, danger, we pluck this flower, safety. [King Henry IV, Part I, Act II, Scene III.]

(Although Hotspur is ultimately killed in battle by the Prince of Wales, he makes a timeless point.)

Wildavsky's argument — although he doesn't say so explicitly — is built around two of the most seminal concepts in the social sciences: 'unanticipated consequences' and 'opportunity costs'. The sociologist Robert K. Merton's 1936 concept of unanticipated consequences describes the unexpected and (often) unfavourable outcomes of actions taken for other reasons. The economist Friedrich von Wieser's 1876 formulation of opportunity costs describes those costs that we incur because we make one choice rather than another; they are the costs of opportunities foregone. Both concepts have so infiltrated our language and have been generalized so far beyond their original disciplinary use that their origins have been "obliterated by incorporation" into popular and scientific discourse — to use another of Merton's concepts.

Stated simply, Wildavsky's thesis is that

many of the measures we adopt in our search for safety have the unintended consequence of creating greater danger. At the same time, these unwise actions have caused us to incur opportunity costs: we could have done other things which would have increased our safety.

On one level this thesis is rational and tactical, but beneath it lies an ideological position, which constitutes, in true Shakespearean tradition, a play within the

Safety in numbers — opponents of a proposed nuclear waste disposal site at Bradwell in Essex.

play. There is a reasonably strong correlation in the industrialized countries between political ideology and attitudes towards technology. The left tends to fear the products of capitalist industry — it seeks government regulations to protect the consumer; it prefers 'soft' energy to 'hard'; and above all it is distrustful of nuclear energy. The right, on the other hand, tends to think that market competition will eliminate dangerous products — it favours deregulation of industry; and it prefers nuclear energy to what Wildavsky refers to as "misplaced nostalgia", by which he means such technologies as wood stoves and solar energy. (There are of course exceptions to this polarization, such as the aristocratic support of environment and wildlife preservation movements.) Wildavsky's sympathies are with the right, as he freely admits, and he aligns himself with the "cornucopians" rather than with the "catastrophists". The extent to which his ideology colours his analysis is a question I think each reader must answer individually.

Wildavsky describes many examples of the unanticipated consequences of the search for safety, ways in which the search actually increases risks. He examines the relationship between adding safety

devices and procedures and achieving greater safety, and finds the relationship more likely to be nonlinear than linear; that is, adding more safety devices and procedures does not lead to a corresponding increase in safety. The accident at Chernobyl, he reports, was the result of on-line testing carried out to increase reactor safety. Similarly, he asserts that the pattern of massive spying in the US Navy that was uncovered in 1985 went undetected for nearly two decades because of the over-classification of many documents. This example leads him to formulate Gresham's law of risk: false alarms drive out true ones. He denounces the courts in the United States for invoking the concept of 'crashworthiness' in pronouncing upon cases involving automobile accidents. A truly crash-worthy car, a study determined, would

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

weigh about 50 per cent more than present cars, and thus do so much damage in collisions with smaller cars that the national death toll would climb until all small cars had been replaced.

Many examples of the unanticipated consequences he cites are from medicine. A distinguished Soviet scientist suffered a pulmonary embolism while mountain climbing at 20,000 feet. As he was eminent, he was evacuated to a Moscow hospital by helicopter; if he had been left near the accident, physicians familiar with this particular illness might have saved him. Similarly, "hospitals make people sick, exercise can hurt you, herb tea is laden with carcinogens" (p. 37). Writing before the current dispute over the relationship between Accutane and birth defects, Wildavsky supports the view that no medicine can be foolproof. He cites as a classic case of opportunity benefits foregone what he calls the "drug lag", the requirement of the US Federal Drug Administration — enacted in the wake of the thalidomide tragedy — that new drugs be proven 'safe' before they can be marketed.

Although Wildavsky has a tendency to create slogans and to shoot from the hip, he has a central point that is worthy of attention. Our emphasis upon what he

derisively calls "anticipation" leads us, because of our ignorance of future events, to bear opportunity costs that are unnecessary. His proposed solution is that we should spend our efforts not in anticipating what will go wrong but in creating resilience in the system.

Many of his solutions borrow from ecology. Wildavsky points out that living systems cannot anticipate specific dangers. Rather, they develop resilience, "the capacity to cope with unanticipated dangers after they become manifest, learning to bounce back" (p. 77). Other solutions are based upon neoclassical economics. Wildavsky has great faith in entrepreneurial activity, in trial-and-error risk-taking (rather than risk aversion) and in market competition, and at the heart of

his analysis is the proposition that the high-risk, entrepreneurial market system is the main reason for the rapid increase in health, safety and longevity. Competition increases wealth and distributes the discovery process throughout society, which increases both resilience and the rate of innovation. "Whether society should mainly seek to increase its ability to respond to unexpected dangers by increasing its resilience, or whether it should seek to anticipate dangers to prevent them from doing harm, is what the risk debate is about" (p. 75). Wildavsky makes a strong if polemical case for resilience over anticipation. □

David L. Sills is an Executive Associate at the Social Science Research Council, 605 Third Avenue, New York, New York 10158, USA.

Current questions

Peter D. Killworth

Principles of Ocean Physics. By J.R. Apel. Academic: 1987. Pp.634. Hbk £38, \$76; pbk £19, \$39.

WRITING a textbook is a labour of love. The amount of material to be covered, especially in oceanography, must be daunting to any author. It is then refreshing to find that Apel has not restricted his treatment to the geographical description plus survey of geophysical fluid dynamics that is traditional in the literature. Instead, he includes coverage of acoustics, electromagnetics and optics, largely as seen from a physicist's viewpoint.

This is a praiseworthy approach, one not to be found in other textbooks, but such wide subject matter presents problems. It seems that in order to keep the text to a reasonable length, Apel has had to limit the depth of his treatment, and refer the reader as appropriate to other, more detailed sources. Sadly, that has meant that the book is not as self-contained as ideally it should have been.

There are also other flaws which detract from its effectiveness. Forward references abound, often leaping over several hundred pages, and the style of writing leaves something to be desired. Many sentences are over-long; I often had to reread whole paragraphs in order to understand them. Symbols change their meaning and acquire new dependences on other dimensions, not always with warning. Units do the same, switching from cgs to mks several times on a page, or between diagrams. For example, Fig. 9.7 is in MJ day⁻¹m⁻², and Fig. 2.5 in W m⁻²; Apel provides a conversion factor, for readers with calculators, but a redrawn diagram would have been more useful. I also question the correctness of placing units after a symbolic algebraic equation.

The unit W(m²Km⁻¹)⁻¹, found on p. 141, although intended to help the reader, must surely do the opposite.

The index, a cornerstone of any textbook, is worse than useless. "Density of seawater" has one reference! "Debye absorption" leads the reader to "something like Debye absorption (see Chapter 8)"; after searching therein, a reference to "dipole moment (see Section 4.1)" is found — but dipole moment is not mentioned in Section 4.1. Should the student forget the definition of, say, zenith angle, the index will not help him. I collected a full page of omissions of this nature, and there were many more.

Despite the lengthy errata, many mistakes remain. For example, equation 7.14 defines the modulus of a vector to be a complex number; the left-hand-side of equation 8.55 is real, the right-hand-side imaginary; and the Ekman number as defined in Appendix 5 is identically unity.

Geophysical fluid dynamics is made easier because certain terms are small and can be omitted ('scaling'). Almost no mention of this is made, making the treatment hard to follow. The placing of material is odd, too: steady motions are dealt with in a chapter entitled "Waves and Tides", and quasigeostrophic motion appears to be defined as a small perturbation to an eastward flow. It is hard to find out from the text how large typical values of quantities are: how does the reader interpret a 'large dipole moment' of 1.84 debye without some feel for the size of an ordinary one?

It grieves me not to be able to be more enthusiastic about this book. I applaud the author's motives, but the result is confusing and littered with errors. It will be of some use to those seeking a feel for physical oceanography, but new graduate students should look elsewhere. □

Peter D. Killworth is a Principal Scientific Officer at the Deacon Laboratory, Institute of Oceanographic Sciences, Wormley, Godalming, Surrey GU8 5UB, UK.

Treatment of the future

David Weatherall

Human Gene Therapy. By Eve K. Nichols. Harvard University Press: 1988. Hbk \$22.95, £18.50; pbk \$9.95, £7.95.

IN A FIELD that is moving so rapidly that anything other than overnight success is unthinkable, human gene therapy seems to have been just around the corner for an embarrassingly long time. But having been bitten once by a premature attempt, clinicians will want to ensure that several prerequisites are fulfilled before their patients are subjected again to somatic gene therapy — that is, the transfer of genes into cells other than those of the germ line in order to cure an inherited disease.

First, to counsel prospective patients or their families it must be possible to give an accurate account of the clinical course of the illness and its response to more conventional therapy; genetic diseases vary in their severity, and even when the underlying molecular lesions are known the reasons for their clinical heterogeneity often remain unexplained. Second, we must be able to find the appropriate gene and define its main regulatory regions. Third, we have to identify and isolate the appropriate target cells and develop efficient and safe vectors with which to transfuse them. Finally, there must be clear evidence from animal experiments that the inserted gene will function adequately, that the recipient cell population will have a reasonably long life-span, and that the gene we have transferred will produce no deleterious effects in its new home. *Human Gene Therapy* shows just how far we are from meeting these requirements.

In October 1986, the Institute of Medicine of the National Academy of Sciences held a conference on gene therapy and the organizers invited Eve Nichols to write an account of the meeting for non-experts. She has succeeded remarkably well in this daunting task, and has set out both the scientific background and ethical problems with great clarity.

After introductory chapters that describe how recombinant DNA technology has been applied to the study of human disease, the remainder of the book is devoted to an explanation of the use of retroviral vectors for gene transfer. It concludes with a short section on ethical and economic issues and how, in the United States, federal regulation of programmes for gene therapy is being organized. Eve Nichols has the unusual knack of being able to simplify complex matters without compromising scientific accuracy. She has

been helped by some clear sketches that explain the mysteries of retrovirology.

Because the meeting on which this book is based was confined to the retroviral approach to gene transfer, *Human Gene Therapy* does not provide a comprehensive picture of the field. For example there is no discussion of the possibility of gene replacement by site-directed recombination or by the reactivation of the fetal globin genes to treat the more severe haemoglobin disorders. And some of the ethical issues are dealt with in rather a parochial way. But, with these minor reservations, the book is an excellent introduction for newcomers to this biomedical minefield.

Although somatic gene therapy is no different in principle to organ transplantation, it remains a highly emotive topic. Many clinicians are concerned that, in their haste to be first, workers in this field will cut corners, both scientific and ethical. Anyone who has such worries should read this book. Its clear and critical account of the complexities of the problem, both scientific and ethical, will do much to reassure them that researchers in the United States are putting their house

in order. The recent guidelines on gene therapy research issued by the European Research Councils should be equally comforting.

It is particularly reassuring to read that work is being confined to somatic gene therapy. The alternative approach, germ-line therapy, in which genes are inserted into fertilized eggs and hence passed on to future generations, is a non-starter; it should remain so. Provided that, with adequate regulation, research on the very early human conceptus is allowed to continue it should be possible to learn how to identify a genetic disease in fertilized ova and to return to the mother only those which do not carry the harmful gene. But although prenatal diagnosis of genetic disease will reduce the number of affected births, many disorders will continue to go undetected. This is why somatic gene therapy will play an important role in clinical practice in the future. But judged by this book, this *tour de force* of medical science will remain just round the corner for a while yet. □

Sir David Weatherall is Nuffield Professor of Clinical Medicine in the University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Natural goodness

Vernon Reynolds

The Waning of Humaneness. By Konrad Lorenz. Translated by Robert Warren Kickert. *Unwin Hyman: 1988. Pp. 250. £12.95.*

PUBLISHED in German in 1983, only now has this profound book been translated for an English-speaking public. The Lorenz who speaks here is the same Lorenz from the days of *On Aggression*. I recall an ethology congress in the 1960s when we, the delegates, wore flowers in our button-holes for the great man's address, which was on "The Decline of Youth" or some such topic. The same theme occurs in this new book, 20 years on. Youth is not so much blamed as offered condolences for its boredom and deep distrust of the older generation.

The old thinking, which Lorenz now calls "evolutionary epistemology", has been extended and deepened. It is now more coherent and philosophically rooted than it was then. We human beings are viewed as products of nature, the potentials and limitations of the human mind arising from our animal background. Its potentials include joy and wonder and creativity; its limitations are a poverty of comprehension in large-scale society and a willingness to follow demagogues and those offering false hopes. These limitations are dubbed "malfunctions of once meaningful behaviour patterns".

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Konrad Lorenz — opposing urbanization.

The evolutionary theory of knowledge that Lorenz here espouses is a human-based theory in a very literal sense. It is not founded on mathematical logic nor on reason nor on any theory of teleonomy or of inevitability. It begins and ends with a particular evolved species — the human being — and with the psycho-physiological properties of that species. In a remarkable insight, Lorenz discovers that, from this premise, human emotional feelings and sense-data are "valid", as indeed are the emotions and perceptions of each and every species. They are indeed more valid than the truths declared by mere verbal rationality, and far more so than the very suspect values of modern technocratic society. Empirical science may lay claim to valid truths, but the truths we

perceive and feel are valid too.

From this insight Lorenz assembles his naturalistic philosophy. Man is basically good; civilization distorts the evolved feelings and perceptions that arise in each of us. The *a priori* values, the natural goods and evils, have become perverted by our Western culture, and modern man is now totally confused about what is good and what is evil. In our huge societies, a small power élite controls the mass of the population; things have gone psychologically wrong on both sides of the divide. Big business in particular is blamed for man's loss of humaneness; small enterprises can be humane but they are powerless against the giants.

Worst of all are the urbanized. They "come in contact almost exclusively with only non-living and, above all, with man-made things, and they have learned to come to terms with almost all of them". However, "they treat everything living... with sheer, unbelievable shortsightedness... they regard everything as makable... These epistemologically mistaken paths of human thought and reasoning have irremediable consequences" (pp. 166–167).

Enough. What is to be made of Lorenz's philosophy? Two criticisms of it seem in order. First, it claims universal validity. What is small is good, what is orderly is good, what is natural for man to think, perceive and do is good. But Lorenz himself is the arbiter of the natural, and that surely won't do. He would claim that nature, ethology, other species are the arbiters. But this is not so; there has been a new wave of sociobiological concepts, postdating Lorenz, and these generate quite different ideas of the natural. What is natural is not self-evident even in the animal world, let alone the human.

Secondly, despite his evident efforts to place his ideas in the context of the work of others, Lorenz has concentrated too narrowly on the ideas of a few German philosophers (Kant, Gehlen, Planck and Spengler receive mention). His positive evaluation of the natural over the urban and man-made certainly goes back to Virgil's *Georgics*, and although his is a scientific philosophy, based on the theory of evolution and the findings of empirical research, his ultimate assumption about the validity of emotional and sense data is only scientific in the widest sense — indeed he attacks conventional reductionist science for its inadequacy and bigotry. This must place him in the camp of the natural philosophers generally, for they mostly eschew science in the narrow sense. Wordsworth's pantheism is not far from Lorenz's evolutionary epistemology when the chips are down. □

Vernon Reynolds is a Lecturer in the Department of Biological Anthropology, University of Oxford, 58 Banbury Road, Oxford OX2 6QS, UK.

Reflections of the past

Ian Stewart

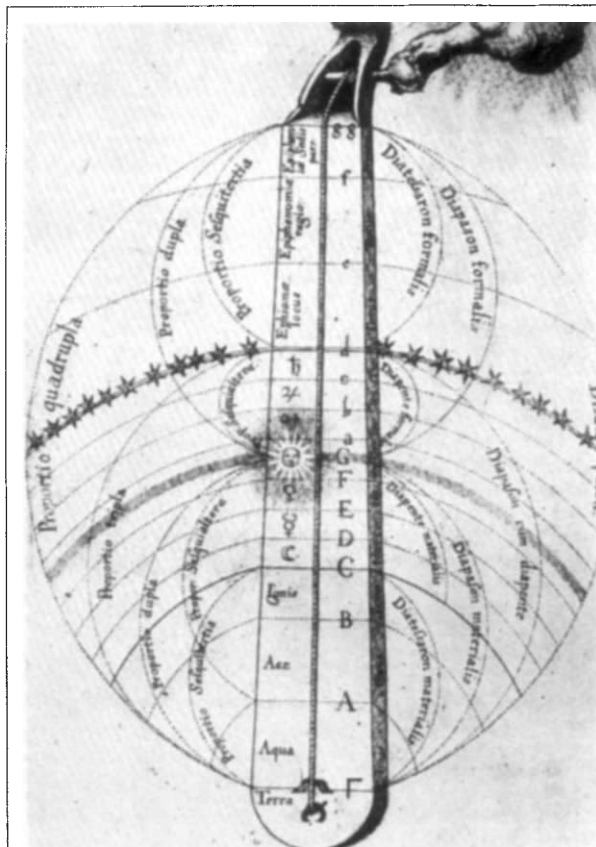
Felix Klein and Sophus Lie: Evolution of the Idea of Symmetry in the Nineteenth Century. By I. M. Yaglom. Translated by Sergei Sossinsky. Birkhäuser: 1987. Pp.237. \$40.

KLEIN and Lie, two giants of the nineteenth century, were between them responsible for deep-seated changes in the way mathematicians think about geometry. For Euclid there was only one geometry, *the* geometry of physical space. By the time of Bolyai, Gauss and Lobachevskii, Euclid's had become just one of three distinct possible kinds of geometry. Soon three had become dozens; then, in the hands of Bernhard Reimann, an infinity. To control this proliferation, some general organizing principles were needed. They were supplied by Klein and developed into a deep and marvellous theory by Lie. Before them, geometry was the study of shapes in space; after them it was the study of invariants of transformation groups.

For example, consider the concept of an isosceles triangle. For Euclid, this is a figure made up from three straight lines, two of which have the same length. Its other properties, such as two angles being equal, are proved by logical arguments based on a small number of fundamental axioms. To Klein, on the other hand, it is not the triangle that is important, but the rigid motions of the plane that preserve its shape and size: rotations, reflections, translations. The equality of two sides is a consequence of the triangle's mirror symmetry about a suitable line; and so is the equality of the corresponding angles. Complicated exercises in logic become transparent applications of symmetry.

Thus the concepts of Euclid's geometry are reformulated as quantities, such as lengths and angles, that remain unchanged under some collection of transformations. These transformations are said to form a group, which means that the result of applying any two of them in turn is just another permitted transformation. Distinct geometries become distinct groups: the surface features are suppressed and the underlying abstract structure is brought to the fore.

Felix Klein and Sophus Lie is an attempt to describe the mathematics involved in this development, and the people behind it. This is an ambitious task, and to some extent the book falls between two stools. On the one hand, it is an authoritative account of an extremely important series of developments that changed the face of mathematics, and most of it is accessible



World harmony — the universal monochord of Robert Fludd, which identifies 'musical' ratios among the radii of various spheres in the Elemental, Celestial and Emphyrean regions. The theories of Fludd (1574–1637) on harmonics and cosmology were inevitably compared with those of Johannes Kepler (1571–1630). In his *Harmonices Mundi Libri V*, Kepler points out that Fludd's work differs from his own as "a Practitioner differs from a Theoretician, for where he deals with instruments I inquire into the causes of things or consonances". The illustration is taken from Kepler's *Geometrical Cosmology* by J. V. Field, published by Athlone Press/Chicago University Press, price £32, \$37.50.

to a non-specialist. On the other hand, the choice of material is more than a little idiosyncratic, the contents do not adequately reflect the title or the subtitle, and there are passages that are far too technical for the non-specialist. It is also rather fragmented.

Taken at face value, the book deals with the discovery of group theory and non-euclidean geometries. It takes up the story around the time of Evariste Galois, who tackled the general question "When is a polynomial equation soluble?". There is a brief excursion into the geometric works of Darboux and Plücker, after which the story drops back a century or so to the early days of projective geometry. The tangled tale of Gauss, Lobachevskii and Bolyai leads, via non-euclidean geometry, into hypercomplex numbers and the foundations of vector algebra. There follows a chapter on Lie and another on Klein's *Erlangen programme* — effectively the statement that geometry is group theory. A final chapter adds some biographical details.

This occupies the first 137 pages. There follow another hundred pages of notes, dotting the i's, crossing the t's and filling in the technicalities. The history is sound, the mathematics important, the exposition accurate — but the overall effect is unsatisfactory. For example the chapter that is supposed to describe Lie's theory of continuous groups (now called Lie groups) is in fact devoted to the formal properties of associative algebras, leading up to the concept of a Lie algebra. Two

pages of small print are spent describing the classical families of simple Lie groups in terms that require far more mathematical maturity than the surrounding text. The small print is intended as a warning, but footnotes are surely a more appropriate place for such material. A reader who does not already know will get little idea of what a Lie group is; one searches in vain for a clear indication even of what the word 'continuous' is supposed to mean in this context. And accessible examples that emphasize the role of continuity, such as the rotation group in three-dimensional space, are absent.

The reader might also be forgiven for emerging with the belief that the main reason why Lie groups are central to today's mathematics is that they generalize Galois Theory to differential equations, when in fact this is of relatively minor importance compared to their role in topology, dynamics and indeed group theory itself. There is also remarkably little said, anywhere, about symmetry — the subtitle notwithstanding.

The notes may well be the book's saving grace; one can learn a lot from them. They include a fascinating discussion of mathematical anti-Semitism, a page on Cayley's 'octaves' (or octonions as they tend to be called now) and a brief history of vector calculus. Maybe that is the point — this is not really a book, but a set of notes for a book. □

Ian Stewart is a Reader in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Could cruise missiles stop START?

Jeremy Leggett

Sea-launched cruise missiles are outside the scope of the US-Soviet treaty on intermediate-range nuclear forces. The consequences for the next step in arms control will be far-reaching.

WITHIN three years, cruise missiles are due to be removed from the European landmass for good, along with the other weapons involved in the US-Soviet INF agreement, signed at the Washington summit in December last year. But while these reductions take place, nuclear cruise missiles of the same range and explosive power will continue to be deployed at sea, as they have been since 1984.

Nuclear-tipped cruise missiles on ships and submarines are more difficult to detect than cruise-missile deployments in rural Europe. Yet they must be limited if a strategic arms reduction treaty (START) is to be signed. Since the Moscow summit two months ago, the chief US negotiator has made it clear that the difficulty of verifying limitations on naval deployments of cruise is the main obstacle to achieving a START treaty¹, and with it cuts of up to about 40 per cent in the superpowers' long-range warheads. Here I describe how that situation came about, and assess the consequences of continuing failure to find a technical or political solution to the verification of sea-launched cruise missiles (SLCMs).

The US Navy began deploying nuclear land-attack SLCMs, called Tomahawks, in 1984. By 1994 a total of 758 nuclear Tomahawks are due to be in place on US ships and submarines. By that time, 197 ships and submarines — fully one-third of the US fleet — is due to carry them. The US Navy also plans to deploy more than 3,000 identical missiles with conventional warheads; these missiles are 'dual-capable', and can be converted to carry nuclear weapons. More than 250 nuclear Tomahawks have been installed already. The sea-launched version of the nuclear Tomahawk is virtually identical to the ground-launched cruise missile of Greenham fame: the maximum range of both is 2,500 km, and both deliver 150 kiloton warheads (equivalent to roughly ten Hiroshimas). One difference is that land-attack SLCMs count as long-range, or strategic, systems — they can be carried, like air-launched cruise missiles, most of the way to the target. Another is that deployment at sea takes place further from public scrutiny than can deployment on the byways of Europe.

A fairly consistent pattern in the brief history of the nuclear arms race has been that every time the United States introduces a new nuclear weapon technology, the Soviets catch up within a few years. So

it has proved with the SLCMs. The Soviet navy has long carried short-range (up to 450 km) nuclear SLCMs for use against ships. Now it has developed two SLCMs capable, like the Tomahawk, of long-range land-attack. Dubbed 'Tomahawkskis' by the military, the SS-N-21 and SS-NX-24 have a range quoted as

limitations could in principle be negotiated. But in the interim the negotiators have been unable to agree to the means by which this could be done.

Equivocation by the Reagan administration over shipboard inspections has not helped. Because of the ease with which SLCMs can be concealed, part of the verification process might have to involve challenge inspections of the magazines of ships and submarines. In 1986, the US Navy Secretary testified in Congress as follows: "The Navy has made it clear that we are prepared to accept on-site inspection of all of our ships, negotiated on a reciprocal basis, as part of any arms-control agreement" (ref. 4). Cynics might be forgiven for noting that at the time such Soviet military *glasnost* seemed inconceivable. But the world is changing fast. Before the Moscow summit, the chief of the Soviet General Staff, Marshall Akhromeyev, agreed to the opening of Soviet cruise-missile submarines to challenge inspections. The response of the chief US negotiator is illuminating: "That is certainly not acceptable to us. So we will have to find a technical means of verifying" (ref. 1).

Is such a technical fix feasible? In Washington, Mr Gorbachev mentioned to President Reagan that the Soviets had a helicopter-borne device for detecting nuclear weapons. This device, allegedly, not only detects the presence of nuclear weapons aboard ships at a distance of 900 to 1,200 feet, but can also discern the number of warheads, and even the yield. In January of this year, the US negotiators in Geneva asked to see the device, and on 21–23 March, during a visit by Mr Shevardnadze to Washington, the Soviet Union proposed a joint test of the SLCM detector in the Mediterranean.

The United States rejected this idea on the basis that "the Soviets could not explain how this experiment would tell us more than we already know". A senior administration official had told the *Washington Times* on 11 March 1988 that "it seemed to be just a Geiger counter hanging from a helicopter". Such a device, he said, would be easy to fool. On 8 April, a State Department official commented that the Soviet proposals forwarded thus far "simply won't work".

The difficulty of using passive detectors reliably to detect the radiation from fissile material in nuclear warheads at distances of more than a few feet — and the ease of

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Tomahawk takes off, February 1979 — the missile undergoes a successful test off the coast of California.

3,000 km by the Pentagon. The first deployments of the SS-N-21 began at the beginning of this year². The US Defense Intelligence Agency estimates that 1,488 could in principle be deployed on Soviet attack submarines by 1995 (ref. 3). The Tomahawkskis represent a little-discussed and probably little-appreciated threat to the security of Europe.

Despite their copy-cat activities, before the Reykjavik summit in October 1986 the Soviets wanted a total ban on long-range SLCMs. The United States rejected this proposal, and at Reykjavik the two sides decided to defer discussion of the problem. Subsequently, the Soviets offered a concessionary position: to establish a limit of 400 long-range SLCMs per side. The United States professed this to be unverifiable, and maintained that there could be no limitations on SLCMs. At the Washington summit in December 1987, the United States finally accepted that SLCM

Popperfoto

sheltering such warheads from detection — is well understood¹. Nonetheless, despite the physical constraints which would conspire against passive nuclear detectors capable of giving consistently meaningful results, it is surprising that the United States has chosen not to allow the Soviets to at least attempt a demonstration of their device.

Congressional opponents

Congressional opponents of the original SLCM deployments in 1984, fearing an attempt by hawks in the administration to subvert future arms-control prospects by making verification impracticably complicated, amended the 1985 defence budget to require the Pentagon to prepare a report on how arms control on SLCMs could be implemented. The resulting document, obtained by lobbyists under the Freedom of Information Act, was too heavily censored to reveal how the Department of Defense imagined that a technological arms-control lid could be kept on SLCM deployments. The very existence of the report, however, indicates that the verification process — though undoubtedly difficult and intrusive — is possible.

One method for tackling the business of verifying a limitation on small, mobile missiles such as SLCMs is to give the opponent tags to put on his missiles during manufacture, for collection and physical or electronic reading during on-site challenge inspections. But it seems that the United States has put little effort into research on such tagging technology, in readiness for a possible change in stance over on-ship inspections during future negotiations, or into work on other forms of technology for SLCM limitation. The former program leader of treaty verification research at the Lawrence Livermore Laboratory, one of two American institutions where billions of dollars are spent each year on nuclear-weapon design and testing, recently spoke out publicly about federal funding for verification research: "While we have heard a great many calls for on-site inspections and other in-country verification methods", he said, "very little analysis has been done on the specific roles that on-site inspections and other in-country verification activities would play and virtually no research on what technologies and equipment would be used or what measurements would be made" (ref. 6).

The House of Representatives Intelligence Committee, reviewing classified congressional testimony, similarly concluded in November 1987 that: "in its examination of the verification research programs throughout the government, the committee found only scattered efforts to determine the technological challenges that might be faced by US intelligence in future arms-control agreements" (ref. 7).

What is at stake if this superpower impasse over SLCMs continues? There are two outcomes which arms control advocates particularly fear. The first is possible failure to achieve any reductions in strategic warheads, and so take the opportunity offered by the INF treaty to reverse the qualitative nuclear arms race. President Reagan has said of the INF treaty: "previous arms control agreements merely put a ceiling on weapons and even allowed for increases. This [the INF treaty] would reduce the number of nuclear weapons". As things stand, in fact, the superpowers plan to deploy more cruise missiles at sea by the mid-1990s than will be dismantled as a result of the INF deal (442 US, 84 Soviet). These deployments, coupled with NATO's plans to introduce new tactical nuclear weapons to Europe, will make a nonsense of the president's assessment of the INF treaty — unless, that is, he or his successor can deliver the great prize of a START deal.

The second consequence of an impasse over SLCMs is that we will face a world in which threat-assessment is much more difficult, and stability in crises is undermined. Admiral Larry Blose, Director of the Pentagon's Cruise Missiles Project, testified to Congress in March 1987 that SLCM deployments are desirable because they "complicate Soviet planning": "The increased strike range... presents the Soviets with a formidable threat from a 360 degree axis against which they have no reliable defense" (ref. 8). As a result, Blose believes, "Tomahawk will conserve valuable resources and save lives". Nuclear Tomahawks "allow us to hold at risk over 90% of Warsaw Pact military and leadership targets and key military facilities in the Soviet North Pacific".

The US Navy's view is that this situation enhances deterrence. But, in Blose's own words, "should deterrence fail", the Tomahawks increase "the range of escalation options available to the National Command Authority without resorting to strategic systems". "Escalation options" is a euphemism for nuclear war-fighting: Blose and his colleagues believe that if the worst comes to the worst, the US Navy will be able to bring 150 kiloton warheads down on the Soviet homeland without necessarily causing the Soviets to resort to their strategic arsenal.

Blose concluded his presentation to Congress by professing that the US Navy is "about to revolutionize naval warfare in a most cost effective manner".

Even as the Moscow summit was announced, the Soviet navy had begun to expand its already considerable seaborne nuclear capability in response to the threat of continuing Tomahawk deployments. The SS-N-21 Tomahawski deployments are beginning at a time when NATO has embarked on a controversial naval doctrine. This, the Maritime Strategy,

involves forward movement of naval forces in a time of superpower tension into waters which the Soviet Union regards as its own, coupled with openly avowed efforts to find and sink Soviet ballistic missile submarines when the first shot of a conventional war is fired. More individual commanders — on both sides — will have the ability in principle to initiate a nuclear exchange. Given that naval nuclear weapons, unlike all others, are without electronic locks (permissive action links), the scope for nuclear war to start by accident or miscalculation increases.

Additionally, both sides will find it more difficult to judge the capabilities of the other, and so fears of pre-emption will grow. This will be especially so when stealth technologies are introduced to cruise missiles, as the United States is intending to do in the years to come. The pressure to 'go first' if the superpowers stumble into another Cuban missiles crisis — much analysed in the detached and arcane literature of nuclear strategy — will build.

Amid all the uncertainty and long-standing deadlock over SLCMs has recently come a remarkable intervention. Ambassador Paul Nitze, President Reagan's veteran arms-control advisor, a man who served as US Navy Secretary in the 1960s, proposed in April the maritime equivalent of the 'zero option'. Nitze's suggestion to the Reagan administration was that both sides should give up *all* nuclear SLCMs, plus naval tactical nuclear weapons (depth charges, torpedoes and the like).

Military glasnost

As any arms-control student will know, it is far easier to verify whether an opponent is sticking to a level of zero than to a particular number of missiles. In the case of nuclear SLCMs, there are also reasons of stability to consider when weighing Nitze's remarkable proposal. But in any event — nautical zero option or no nautical zero option — the US Navy will have to catch up with the Soviets in the question of military *glasnost*. □

1. Kampelman, M. United States Information Service, Embassy of the USA (June 22, 1988).
2. Mack, A. *Bull. Atomic Sci.* **44**, 7-9 (1988).
3. Rubin, J.P. *Arms Control Today* (2-11 April, 1986).
4. Lehman, J. *Testimony to the Subcommittee on the Department of Defense, Committee on Appropriations* (House of Representatives, US Congress, 1986).
5. Gspooner, A., in *Nuclear Disengagement in Europe* (eds Lodgaard, S. & Marek, T.) 209-219 (Taylor & Francis, London, 1983).
6. Nordyke, M. Lecture at American Association for the Advancement of Science Annual Meeting (1987).
7. Permanent Select Committee on Intelligence, US House of Representatives. Report 100-450 (US Congress, 1987).
8. Blose, L. *Testimony before the Procurement and Military Nuclear Systems Subcommittee, Armed Services Committee, US House of Representatives* (12 March, 1987).

Jeremy Leggett is a Reader in the Department of Geology, Imperial College of Science and Technology, University of London, London SW7 2AZ, UK, and Director of the UK Verification Technology Information Centre (VERTIC).

Phase transitions of stored laser-cooled ions

R. Blümel, J. M. Chen, E. Peik, W. Quint, W. Schleich, Y. R. Shen* & H. Walther

Max-Planck-Institut für Quantenoptik, D-8046 Garching, FRG

Single ions in miniature traps can be imaged by using laser light to stimulate fluorescence radiation. At the same time, radiation pressure can be used to bring them nearly to rest. When a small number of ions are trapped, phase transitions can be observed between a chaotic cloud and an ordered crystalline structure, depending on the degree of laser cooling.

THE behaviour of charged particles in an external potential has been discussed frequently in different contexts. Examples include the Thomson 'raisin model'¹ and strong electrolyte solutions, treated by Debye and Hückel². The first quantum treatment of the problem was presented by Wigner³ who proposed a new phase of the degenerate electron gas, the Wigner crystal.

Experimental studies of charged particles in an external potential reached a new stage after the invention of ion-traps^{4,5}. The first observation of crystallized charged particles stored in a Paul-trap was reported for aluminium microparticles (20 μm diameter, 10^5 times the elementary charge) by Wuerker *et al.*⁶, when cooling was achieved by a light background gas. Repeated crystallization of the particles to a regular array and its subsequent melting were obtained by variation of the storage potential and the cooling force. Later, experiments were reported^{7,8} with Be^+ -ions in a Penning-trap^{9,10}. The ions were cooled by radiation pressure¹¹⁻¹³ and it was claimed^{7,8} that plasma parameters^{14,15} (namely the ratio between the potential and kinetic energies of the ions) were high enough for crystallization to be feasible. The observation of the crystallization of Mg^+ ions in a Paul-trap to an ordered structure, its melting and recrystallization (that is, phase transitions) were then reported¹⁶⁻¹⁹. Later, the vibrational frequencies of small ion-clusters of Hg^+ ions were measured²⁰.

Pure electron plasmas have also been produced. They bear the exciting possibility that the quantum regime may be reached at low temperatures²¹⁻²⁴. Recently the problem of ion-crystal formation²⁵⁻²⁹ in strongly coupled Coulomb systems^{30,31} found additional interest in connection with the new ion storage rings with electron- and laser-cooling³²⁻³⁴.

The Paul-trap has so far been directed mainly towards using a single ion^{35,36} for high-resolution spectroscopy³⁷⁻⁴² and as a frequency standard⁴³, which could easily surpass the accuracy of modern Cs atomic clocks⁴⁴.

Here we concentrate mainly on few-body phase transitions that occur for a small number of ions in a Paul-trap between a chaotic cloud and an ordered crystalline structure. We describe our recent experimental and theoretical work on this subject. In particular, we focus on the following questions: what keeps ion-clouds and crystals stable? What are the parameters controlling the transitions between these two phases, and what is the mechanism?

Description of the trap

Central to our experiment is a radio frequency (r.f.) Paul-trap⁴. This device grew out of the work on quadrupole mass filters⁴⁵. Our Paul-trap (see Fig. 1 and ref. 46 for technical details) is larger than most other traps used in laser experiments. The frequency of the r.f. field, f , had a value of $f = 11$ MHz. For the measurement of the secular and vibrational frequencies of ion crystals, an additional a.c.-voltage can be applied between the ring and the two end-cap electrodes.

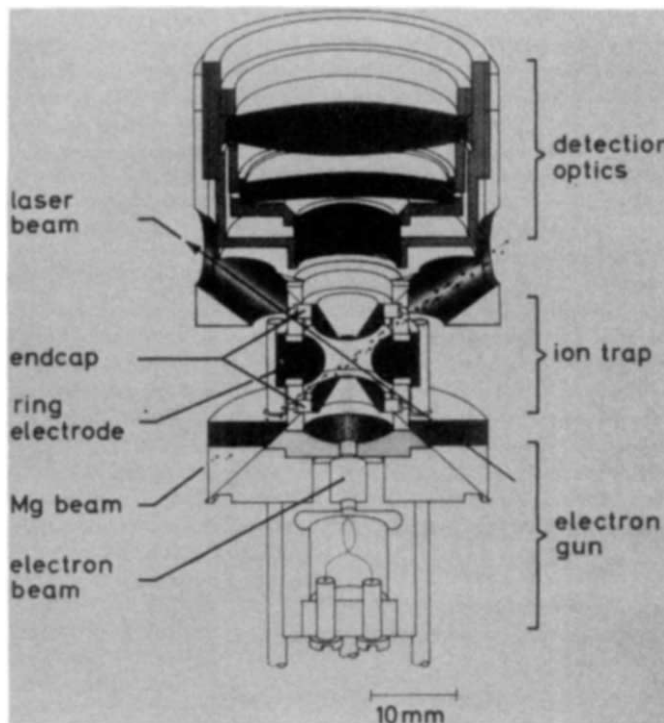


Fig. 1 Sketch of the ion-trap. The fluorescence light is observed through a hole in the upper end-cap. The ionizing electron beam enters through a hole in the lower end-cap. Both holes are covered by a fine molybdenum mesh in order not to distort the trap-potential (for details see ref. 46).

Our trap is mounted inside an ultra-high vacuum chamber and is loaded by a thermal beam of neutral Mg atoms which are ionized close to the trap's centre by an electron beam entering the trap through a small hole in the lower end-cap. The neutral Mg beam and a laser beam are passed through the gaps between cap and ring electrodes. The laser excitation serves two purposes: on one hand it detects the ions by laser-induced fluorescence on the transition $3S_{1/2} \rightarrow 3P_{3/2}$ of Mg^+ , and on the other it cools the ions by radiation pressure¹¹⁻¹³. For this purpose the laser frequency is red-shifted by an amount Δ with respect to the resonance transition (negative detuning). The large size of the trap affords a large solid angle for detecting the fluorescence radiation either with a photo-multiplier or a photon-counting imaging system (Hamamatsu, ARGUS or PIAS). To observe the ions, the cathode of the imaging system was placed in the image plane of the microscope objective attached to the trap (Fig. 1).

Outline of the theory

We now analyse the equations of motion of ions in a Paul-trap. The ions are subjected to essentially four different forces: the

* Permanent address: Department of Physics, University of California, Berkeley, California 94720, USA.

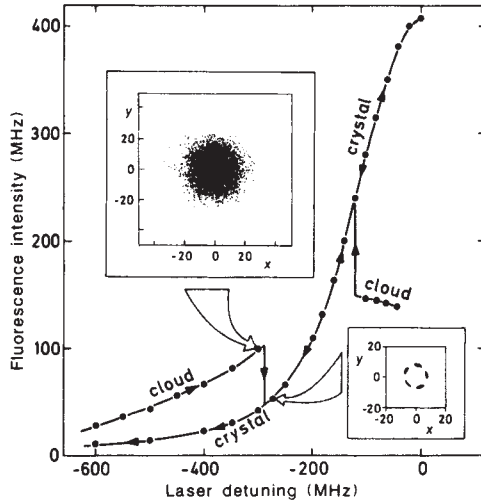


Fig. 2 Theoretical excitation spectrum of five ions as a function of the laser detuning Δ for $U_0 = 0$, $V_0 = 780$ V and $P = 500$ μ W. The black arrows on the fluorescence curves indicate the scanning direction. Two phase transitions (and bistability) are apparent. Insets show the results of three-dimensional molecular dynamics calculations for detunings shortly before and shortly after the jump at $\Delta = -300$ MHz which demonstrate directly the existence of a transition from a cloud phase to a crystalline state. The axes of the insets are in units of μ m. (c.f. Fig. 2 of ref. 18 for experimental spectra.)

force, $\mathbf{F}^{\text{trap}}$, arising from the dynamic r.f. trapping field, the Coulomb interaction between the ions, $\mathbf{F}^{\text{Coul}}$, the laser cooling force, $\mathbf{F}^{\text{laser}}$, and a random force, $\mathbf{F}^{\text{rand}}$, arising from the spontaneously emitted photons.

We now discuss these forces in detail. The potential generated by the r.f. voltage applied to the hyperbolic electrodes of the trap is given by^{4-6,9}

$$\varphi(\mathbf{r}, t) = g(t)[x^2 + y^2 - 2z^2]; \quad g(t) = \frac{U_0 + V_0 \cos(\Omega t)}{r_0^2 + 2z_0^2} \quad (1)$$

where U_0 and V_0 are the d.c. and a.c. components of the voltage, $\Omega = 2\pi f$, and r_0 and $2z_0$ denote the radius of the ring electrode and the separation of the end-caps, respectively. From equation (1), we obtain the trapping force and the resulting equation of motion of a single ion in the Paul-trap

$$\mathbf{F}^{\text{trap}} = m\ddot{\mathbf{r}} = -e\nabla\varphi(\mathbf{r}, t) = -2eg(t)[\mathbf{r} - 3ze_z] \quad (2)$$

where m and $\mathbf{r}$ denote the ion's mass and position, and $\mathbf{e}_z$ is a unit vector in z -direction. The equation of motion (2) can be cast into the standard form of a Mathieu equation⁴⁷ which, depending on the voltages U_0 and V_0 , shows domains of stability and instability. Generally, the ion's motion in the trap is composed of a superposition of two motions: (1) the secular motion corresponding to the harmonic vibration of the ion in the pseudo-potential of the trap^{4-6,9} (frequencies $\omega_x = \omega_y$ and ω_z) and (2) the micro-motion^{4-6,9}, which arises from the $\cos(\Omega t)$ driving term. Generalizing equation (2) to the case of n particles, the resulting trapping force reads

$$\mathbf{F}_i^{\text{trap}} = -2eg(t)[\mathbf{r}_i - 3z_i\mathbf{e}_z] \quad (3)$$

The Coulomb interaction between the particles results in the force

$$\mathbf{F}_i^{\text{Coul}} = \frac{e^2}{4\pi\epsilon_0} \sum_{m \neq i} \frac{\mathbf{r}_i - \mathbf{r}_m}{|\mathbf{r}_i - \mathbf{r}_m|^3}. \quad (4)$$

In the process of laser cooling, every scattered photon changes the momentum of an ion on average by an amount $\hbar\mathbf{k}$, resulting in the laser cooling force⁴⁸, $\mathbf{F}_i^{\text{laser}}$, acting on the i -th ion

$$\mathbf{F}_i^{\text{laser}} = \hbar\mathbf{k}R[\mathbf{r}_i(t); \mathbf{r}_i(t)] \quad (5)$$

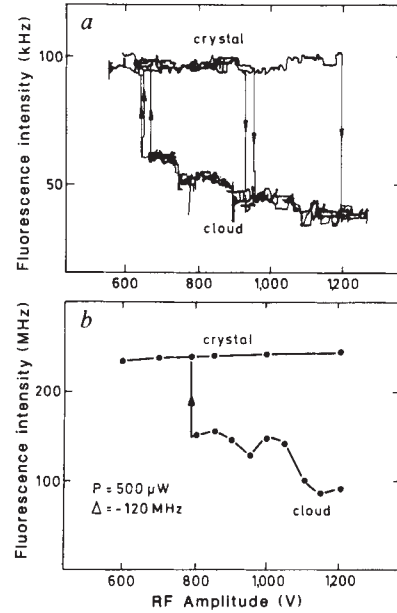


Fig. 3 Comparison between *a*, experimental and *b*, theoretical fluorescence intensity of five ions for identical trap parameters ($U_0 = 0$), as a function of r.f. amplitude. The experimental fluorescence intensities are not corrected for detection efficiency.

The photon scattering rate, R , used for our calculations takes saturation of the atoms and the gaussian profile of the laser beam into account. The number, $N_i(l)$, of spontaneously emitted photons from the i -th ion during the l -th cycle of period $T = 1/f$ is given by

$$N_i(l) = \int_{(l-1)T}^{lT} R[\mathbf{r}_i(t); \mathbf{r}_i(t)] dt \quad (6)$$

In our simulations, we emit these N_i spontaneous photons, one by one, at the end of each r.f. cycle, rather than emitting them, one by one, according to the appropriate photon-statistics during the cycle. The direction $\hat{\mathbf{u}}_j$, of the j -th emitted photon is chosen at random, weighted by a $\cos^2$ -distribution with respect to the laser polarization axis. This way, only the linearly polarized spontaneously emitted photons have been simulated in our calculations. The circularly polarized ones, occurring less frequently with a $\sin^2$ -distribution, have been neglected. The random force, $\mathbf{F}_i^{\text{rand}}$, is thus

$$\mathbf{F}_i^{\text{rand}}(t) = \hbar\mathbf{k} \sum_{l=-\infty}^{\infty} \delta(t - lT) \sum_{j=1}^{N_i(l)} \hat{\mathbf{u}}_j \quad (7)$$

Altogether, the motion of the i -th particle is governed by the sum of the above forces, that is,

$$m\ddot{\mathbf{r}}_i = \mathbf{F}_i^{\text{trap}} + \mathbf{F}_i^{\text{Coul}} + \mathbf{F}_i^{\text{laser}} + \mathbf{F}_i^{\text{rand}} \quad (8)$$

Given appropriate initial conditions, the equations of motion (8) are integrated forward in time by a standard Runge-Kutta⁴⁹ procedure to obtain the trajectories of the n -ion system. These trajectories provide all the information necessary to find the quantities of interest, such as the (total) kinetic energy of the ions or the fluorescence intensity (photon counts per second), which is obtained by integrating the scattering rate, R , along the particle trajectories. To eliminate transient effects arising from initial conditions, we integrate the equations until the total kinetic energy has reached a steady state.

Depending on external parameters, such as the laser power P , the detuning Δ and the r.f. voltage V_0 , the solutions of equation (8) can be divided into two classes: in the first the ions form an ordered structure (crystalline phase), and in the second

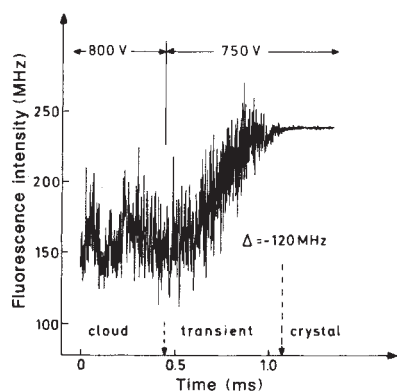


Fig. 4 Time-resolved dynamics of the phase transition of Fig. 3b (with identical trap parameters) extracted from three-dimensional molecular dynamics calculations. Horizontal arrows indicate the time periods over which the r.f. voltage, V_0 has the value of 800 V and 750 V respectively.

the ions perform an erratic motion of much larger spatial extension and mean kinetic energy (cloud phase). As a result of the cooling and randomizing effect of the forces F_i^{laser} and F_i^{rand} , the equilibrium energy is independent of the particular choice of initial conditions. For those trap parameters, however, where both phases occur and are stable (bistability), the phase space of initial conditions is divided into two domains representing the basins of attraction for the crystalline phase and the cloud phase, respectively.

Experimental and theoretical results

In this section we discuss our few-ion phase transition experiments in detail and compare them with the results of three-dimensional molecular dynamics simulations of equation (8). Care has been taken in the experiments to ensure that only $^{24}\text{Mg}^+$ ions are in the trap. The effects of the other Mg isotopes, $^{25}\text{Mg}^+$ and $^{26}\text{Mg}^+$, have been studied in detail and will be published elsewhere.

The phase transitions manifest themselves experimentally in abrupt changes in fluorescence intensity when the laser frequency is scanned over critical negative detunings^{18,19}. The transitions between the cloud and the crystalline phase show hysteresis (discussed below). Furthermore, the two phases can be observed directly with the help of a highly sensitive imaging system. It is also possible to identify the crystalline structure by an excitation of vibrational modes using an additional r.f. voltage applied to the trap.

The behaviour of the ions in the trap is governed essentially by three parameters: the trapping voltage V_0 , the laser detuning Δ and the laser power P . Hysteresis in the fluorescence intensity occurs whenever two of these are kept constant and the third is swept. This is shown for the detuning Δ in Fig. 2 and for the r.f. voltage V_0 in Fig. 3. Such hysteresis behaviour can be expected with laser-cooled ions, because the cooling power of the laser is strongly dependent on the details of the velocity distribution of the ions. The jumps from the cloud to the crystal reproduce well in the experiment, as well as in our simulations. The transitions from the crystal to the cloud state, namely the melting of the crystal, occurs only in the experiment (see Fig. 3a), but not in our simulations (see Fig. 3b, for example), where the crystals are stable until the critical voltage corresponding to the instability condition of the Mathieu equation is reached. On the other hand, the experimental data show that the location of the transition from crystal to cloud is not well defined, and the melting of the crystal could be caused by fluctuations in the laser intensity. We note that, depending on the detuning, either the crystal or the cloud can have a higher fluorescence intensity.

What is the timescale on which the cloud-crystal phase transitions occur? To study theoretically the dynamics of a particular

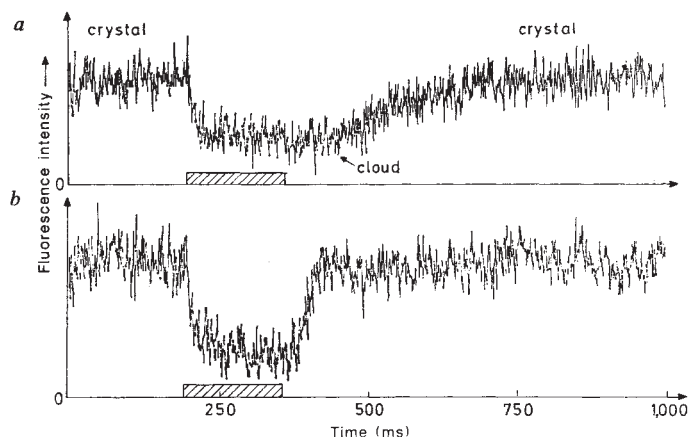


Fig. 5 Time-resolved dynamics of an experimentally obtained phase transition of three ions. $U_0 = 0$, $P = 300 \mu\text{W}$; a, $\Delta = -40 \text{ MHz}$, b, $\Delta = -80 \text{ MHz}$. The r.f. voltage V_0 is suddenly switched from its value 140 V (crystal) to 420 V (cloud), and back to 140 V (crystal). The period over which the r.f. voltage of 420 V is applied is marked by the hatched rectangles. Note that here the number of ions is different from the number in Fig. 4.

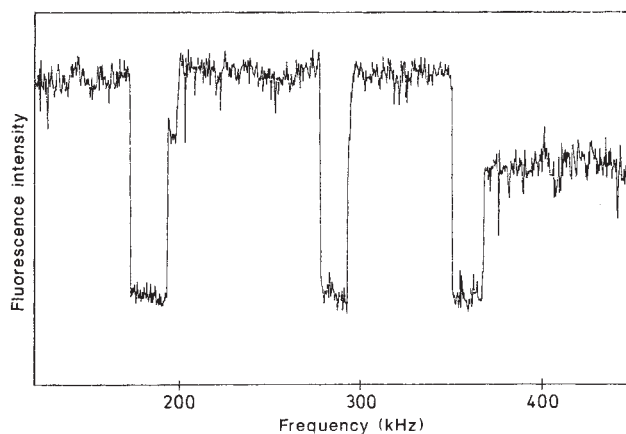


Fig. 6 Spectroscopy of vibrational modes of a two-ion crystal: fluorescence intensity as a function of the frequency of an additionally applied a.c. voltage for $U_0 = 0$, $V_0 = 165 \text{ V}$, $\Delta = -60 \text{ MHz}$ and $P = 500 \mu\text{W}$. The theoretical values of the secular frequencies are $\omega_x/2\pi = \omega_y/2\pi \approx 170 \text{ kHz}$ and $\omega_z/2\pi \approx 340 \text{ kHz}$.

jump, such as the one shown in Fig. 3b, in Fig. 4 we analyse the fluorescence as a function of time, when we suddenly switch the r.f. field from $V_0 = 800 \text{ V}$ (cloud state) to $V_0 = 750 \text{ V}$ (crystal state). After a transient of about 0.5 ms, the value of the fluorescence typical for this cloud state settles to its crystalline value. Experimental time-resolved phase transitions shown in Fig. 5 demonstrate that the transient time strongly depends on the laser detuning. In contrast to Fig. 4, the results in Fig. 5 represent an average over 60 such switching experiments, whereupon the large intensity fluctuations in the cloud state are washed out. The residual fluctuations in Fig. 5 result mainly from counting statistics.

The experimental results described above represent only a small fraction of the entire body of data accumulated. Ion-crystals, clouds and phase transitions of two to approximately 100 ions have been observed.

The most basic ion-crystal—the two-ion crystal—enjoys two kinds of normal modes: (1) three centre-of-mass oscillations with secular frequencies³⁸ $\omega_x = \omega_y$ and ω_z , where the distance between the ions is fixed and, (2) vibrational modes, in which

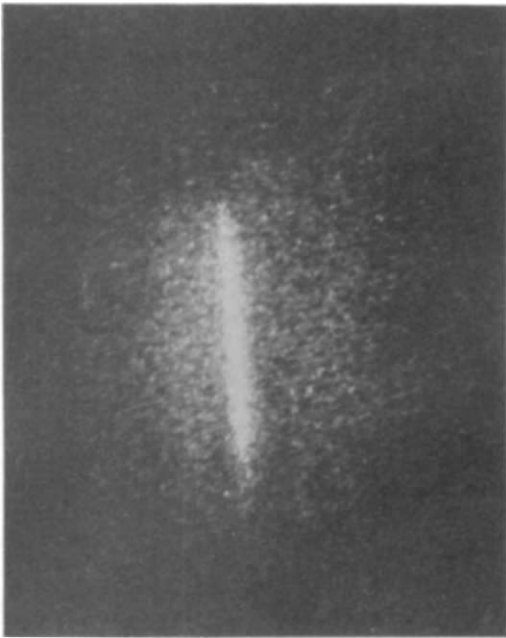


Fig. 7 The large amplitude vibration of a single ion (maximal visible excursion from the trap-centre is $160\ \mu\text{m}$) for the bistable situation $U_0=0$, $V_0=410\ \text{V}$, $\Delta=-100\ \text{MHz}$ and $P=220\ \mu\text{W}$. Increasing the r.f. voltage to $V_0=630\ \text{V}$, the line collapses to a dot. Decreasing V_0 back to its original value, the dot remains stable.

the two ions oscillate against each other with an approximate frequency^{20,29} of $\sqrt{3}\omega_{x,y}$ or $\sqrt{3}\omega_z$, according to the alignment of the crystal, as determined by U_0 . When we apply an additional a.c. voltage to the electrodes of the trap (15 mV), we excite the centre-of-mass motion as well as the oscillation between the two ions (Fig. 6). The resonant excitation of one of these modes results in an additional heating and in a reduction in the fluorescence intensity. The dip in the spectrum at $\sim 170\ \text{kHz}$ thus corresponds to $\omega_{x,y}$, and the one at $\sim 340\ \text{kHz}$ to ω_z . The one in the middle corresponds to the mode in which the ions oscillate against each other. At the end of the recording time, one ion left the trap leaving the fluorescence intensity at half its original value. Figure 6 demonstrates the feasibility of this experimental technique to determine the frequencies of the vibrational modes of ion crystals.

In Fig. 7, a single ion shows a stable large-amplitude vibration orthogonal to the cooling laser beam. According to equation (5), there is only a cooling effect in the k direction. Motion orthogonal to this direction is damped only in a minor way giving to laser-focus effects, contact potentials or trap-imperfections that couple the different degrees of freedom. When we increase the r.f. voltage, the amplitude of vibration reflected in the length of the white line (integrated fluorescence signal) in Fig. 7 is reduced until it suddenly collapses into a dot representing the single ion sitting at the centre of the trap. When we now decrease the voltage again to its original value where the large-amplitude vibration was observed, the ion will still be located at the centre of the trap. This means that there is bistability, even for the case of a single ion. Theoretical calculations confirm this observation and will be published elsewhere.

Discussion

The cooling effect of the laser is balanced by a heating mechanism. Heating of the ions by the r.f. field has been known⁹ since the early days of r.f. traps. Its understanding, however, remained incomplete.

Does the most elementary of all conceivable one-dimensional models for ions in an ion-trap, namely the one-dimensional

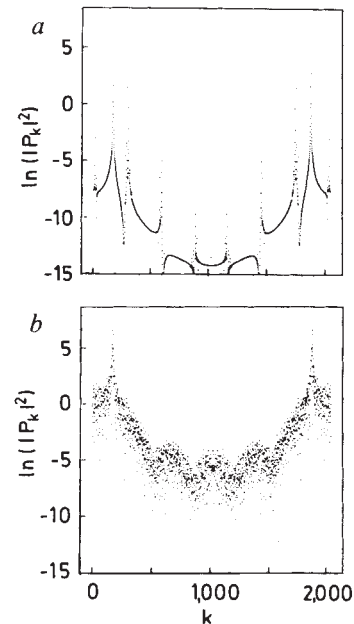


Fig. 8 Theoretical power spectrum of the x coordinate of one of the two ions in a Paul-trap. *a*, Ions started close to the equilibrium position of the corresponding crystalline state. The discrete spikes in the power-spectrum are typical for quasi-periodic motion. *b*, Two ions in the cloud phase: the spectrum is noisy, a characteristic feature of chaotic motion.

version of equation (8), taking into account only $\mathbf{F}^{\text{trap}}$ and the Coulomb interaction force $\mathbf{F}^{\text{Coul}}$, reveal the origin of the r.f. heating mechanism? To answer this question, we have monitored the average energy of two to five ions as a function of time in such a one-dimensional trap. For less than five ions we could not observe any gain in energy over several tens of milliseconds. For five ions, a slow increase in energy was recorded, but it was too small to account for the stability of ion-clouds in the presence of laser cooling. A fast Fourier transform of the positions $x_n(j)$ of the n -th member of the one-dimensional ion-chain taken at time $t=jT$

$$P_k^{(n)} = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_n(j) \exp \left[-i \frac{2\pi}{N} jk \right];$$

$$k = 0, 1, \dots, N-1 \quad (9)$$

($N=2,048$ in the present case) shows a small number of discrete frequencies dominating the spectrum, and thereby qualifies this model as being close to integrable and lacking a heating mechanism.

Two-dimensional ion-traps behave essentially like one-dimensional traps if the motion is restricted to the x - y plane. In the x - z (y - z) plane, however, there is strong heating. To study this observation in more detail, we have calculated the power-spectrum $|P_k|^2$ of the positions of two ions in a two-dimensional trap. We have chosen an initial separation of the ions that was about twice the equilibrium separation of the two-ion crystal, and a set of discrete frequencies appeared, as shown in Fig. 8*a*. In this case, the ions perform a quasi-periodic motion. They are unable to extract energy from the r.f. field and, as a result of the cooling laser, they eventually end up in the crystalline state. Such a power-spectrum characterizes phase space domains that act as basins of attraction for the crystal. But when we choose initial conditions that correspond to typical separations in a cloud state, the spectral power of any one of the two ions is a very noisy, quasi-continuous band which resembles the power spectrum of a turbulent fluid⁵⁰ (see Fig. 8*b*).

There is an interesting connection between the r.f. heating in a Paul-trap and the ionization mechanism of Rydberg atoms. Both are subjected to strong time periodic fields. For a given field strength, microwave frequency and small diameter of the Rydberg atom, the motion of the Rydberg electron is regular and no ionization occurs. An increase of the atom's diameter (by increasing the principle quantum number for example) makes chaotic phase space domains accessible. The motion of the Rydberg electron becomes erratic, resulting in a diffusive gain of energy and ionization of the atom. We encounter an analogous situation for ions in a Paul-trap. If we prepare ions in the vicinity of their equilibrium positions with small initial velocities, they are located in regular regions of phase space. No heating occurs and after a while the cooling laser establishes a stable crystalline configuration. If, however, we prepare an ionic array with small initial velocities, but with a diameter which amounts to about four times the typical dimensions of a crystal, this configuration will be located in a chaotic domain of phase space. Strong heating occurs by a cooperative effect of r.f. driving field and Coulomb non-linearities. Furthermore, our calculations show that the heating rate reduces with the diameter of the clouds. As a result, for very large clouds, the heating stops and the ions will still be confined in the trap. This is confirmed by experiments in which, even in the absence of a cooling laser, large clouds of ions can be stored in a Paul-trap over several hours without being heated out of the trap.

The above considerations and observations suggest a simple explanation for the surprisingly sharp phase transitions found in our experiments. Starting with a large cloud, and increasing the laser cooling power, the cloud reduces in diameter. The increased cooling power of the laser is counter-balanced by the larger heating-rate of the smaller cloud. But as soon as the cloud diameter reaches the critical value separating regular and chaotic domains in phase space, the smallest increase in laser power beyond this value will result in a collapse of the cloud into the crystal state. Diffusion models similar to the ones developed for

Rydberg atoms⁵¹⁻⁵³ could be designed to calculate the energy diffusion coefficient, which would allow for an analytical calculation of the steady-state size of ion-clouds and the exact location of the phase-transition points.

Summary and conclusions

In conclusion, we have shown that ions confined in an ion-trap can be found in two modifications: in a cloud phase and in a crystalline phase. Phase transitions characterized by a discontinuous change in the fluorescence intensity can be induced by appropriately adjusting the r.f. voltage, the detuning and the power of the cooling laser. The transitions show hysteresis as a function of these three essential control parameters. With the help of three-dimensional molecular dynamics calculations, we have reproduced qualitatively all the essential features in recent experiments, such as the size of ion clouds and crystals, as well as transitions between these two phases. The dynamics of the phase transitions were explored experimentally and numerically by time-resolved fluorescence intensity recorded for the transients of specific phase transitions. In detailed studies of one-, two- and three-dimensional simulations of ions in a Paul-trap, we have identified the occurrence of deterministic chaos⁵⁰ in the cloud state as the predominant source of heating. The mechanism of r.f. heating, therefore, relies only on the deterministic features of the system, rather than on the effect of external randomness, such as noise in the amplitude and frequency of the r.f. trapping field, or collisions with the atoms of the background gas.

The static as well as the dynamic features of laser-cooled ions in a Paul-trap is a very promising subject for the study of few-body non-equilibrium phenomena. With the help of present-day powerful imaging systems, crystals, clouds and phase transitions between them can be studied as a function of particle number.

The authors of the present article thank B. Englert, D. Greenberger, A. Schenzle, W. C. Schieve and U. Smilansky for discussions.

Received 20 May; accepted 24 June 1988.

- Föppl, L. *J. reine angew. Math.* **141**, 251-302 (1912).
- Debye, P. & Hückel, E. *Phys. Zeitschr.* **24**, 185-206 (1923).
- Wigner, E. P. *Trans. Faraday Soc.* **34**, 678-685 (1938).
- Paul, W., Osbergerhaus, O. & Fischer, E. *Ein Ionenkäfig* Forschungsberichte des Wirtschafts- und Verkehrsministeriums Nordrhein-Westfalen **415** (1958).
- Fischer, E. *Z. Phys.* **156**, 1-26 (1959).
- Wuerker, R. F., Shelton, H. & Langmuir, R. V. *J. appl. Phys.* **30**, 342-349 (1959).
- Bollinger, J. J. & Wineland, D. J. *Phys. Rev. Lett.* **53**, 348-351 (1984).
- Brewer, L. R., Prestage, J. D., Bollinger, J. J. & Wineland, D. J. in *Strongly Coupled Plasma Physics* (ed. Rogers, F. J. & DeWitt, H. E.) 53-64 (Plenum, New York, 1987).
- Dehmelt, H. G. in *Adv. atom. molec. Phys.* Vol. 3, 53-72 (ed. Bates, D. R. & Estermann, I.) (Academic, New York, 1967).
- Penning, F. M. *Physica* **3**, 873 (1936).
- Wineland, D. J. & Itano, W. M. *Phys. Rev.* **A20**, 1521-1540 (1979).
- Javanainen, J. *J. appl. Phys.* **23**, 175-182 (1980).
- Stenholm, S. *Rev. mod. Phys.* **58**, 699-739 (1986).
- Pollock, E. L. & Hansen, J. P. *Phys. Rev.* **A8**, 3110-3122 (1973).
- Slattery, W. L., Doolen, G. D. & DeWitt, H. E. *Phys. Rev.* **A21**, 2087-2095 (1980).
- Diedrich, F., Krause, J., Rempe, G., Scully, M. O. & Walther, H. in *Laser Spectroscopy VIII* (Springer Series in Optical Sciences **55**, 133-138) (ed. Svanberg, S. & Person, W.) (Springer, Berlin, 1987).
- Diedrich, F., Krause, J., Rempe, G., Scully, M. O. & Walther, H. in *Proc. Fourth Int. Conf. on Multiphoton Processes, Boulder, Colorado, 1987* (ed. Smith, S. & Knight, P.) (Cambridge University Press, New York, in the press).
- Diedrich, F., Peik, E., Chen, J. M., Quint, W. & Walther, H. *Phys. Rev. Lett.* **59**, 2931-2934 (1987).
- Diedrich, F., Peik, E., Chen, J. M., Quint, W. & Walther, H. *Phys. Bl.* **44**, 12-15 (1988).
- Wineland, D. J., Bergquist, J. C., Itano, W. M., Bollinger, J. J. & Manney, C. H. *Phys. Rev. Lett.* **59**, 2935-2938 (1987).
- Malmberg, J. H. & O'Neil, T. M. *Phys. Rev. Lett.* **39**, 1333-1336 (1977).
- Grimes, C. C. & Adams, G. *Phys. Rev. Lett.* **42**, 795-798 (1979).
- Deville, G., Valdes, A., Andrei, E. & Williams, F. B. I. *Phys. Rev. Lett.* **53**, 588-591 (1984).
- Gerhardt, R. *Phys. Bl.* **42**, 23-25 (1986).
- Mostowski, J. & Gajda, M. *Acta phys. polonica* **A67**, 783-802 (1985).
- Baklanov, E. V. & Chebotayev, V. P. *Appl. Phys.* **B39**, 179-181 (1986).
- Casdorff, R. & Blatt, R. *Appl. Phys.* **B45**, 175-182 (1988).
- Javanainen, J. *Phys. Rev. Lett.* **56**, 1798-1801 (1986).
- Javanainen, J. *J. opt. Soc. Am.* **B5**, 73-81 (1988).
- Ichimaru, S. *Rev. mod. Phys.* **54**, 1017-1059 (1982).
- Totsuji, H. in *Strongly Coupled Plasma Physics* (ed. Rogers, F. J. & DeWitt, H. E.) (Plenum, New York, 1987).
- Habs, D. in *Frontiers of Particle Beams* (Springer, New York, 1988).
- Rahman, A. & Schiffer, J. P. *Phys. Rev. Lett.* **57**, 1133-1136 (1986).
- Schiffer, J. P. & Poulsen, O. *Europhys. Lett.* **1**, 55-59 (1986).
- Neuhauser, W., Hohenstatt, M., Toschek, P., Dehmelt, H. *Phys. Rev. Lett.* **41**, 233-236 (1978).
- Neuhauser, W., Hohenstatt, M., Toschek, P., Dehmelt, H. *Phys. Rev.* **A22**, 1137-1140 (1980).
- Dehmelt, H. G. in *Advances in Laser Spectroscopy* (ed. Arecchi, F. T., Strumia, F. & Walther, H.) 153-187 (Plenum, New York, 1983).
- Wineland, D. J., Itano, W. M. & Van Dyck, R. S. in *Adv. atom. molec. Phys.* **19** (ed. Bates, D. R. & Bederson, B.) 135-186 (Academic, New York, 1983).
- Wineland, D. J. & Itano, W. M. *Phys. Lett.* **82A**, 75-78 (1981).
- Nagourney, W., Janik, G. & Dehmelt, H. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 643-646 (1983).
- Janik, G., Nagourney, W. & Dehmelt, H. G. *J. opt. Soc. Am.* **B2**, 1251-1257 (1985).
- Bergquist, J. C., Itano, W. M. & Wineland, D. J. *Phys. Rev.* **A36**, 428-430 (1987).
- Dehmelt, H. G. *IEEE Trans. Instrum. Meas.* **IM-31**, 83-87 (1982).
- Wineland, D. J., Itano, W. M., Bergquist, J. C., Bollinger, J. J. & Hemmati, H. *Prog. Quant. Electr.* **8**, 139-142 (1984).
- Paul, W. & Raether, M. *Z. Phys.* **140**, 262-273 (1955).
- Diedrich, F. & Walther, H. *Phys. Rev.* **58**, 203-206 (1987).
- Whittaker, E. T. & Watson, G. N. *A Course of Modern Analysis* (Cambridge University Press, 1927).
- Loudon, R. *The Quantum Theory of Light* (Clarendon, Oxford, 1986).
- Stoer, J. & Bulirsch, R. *Einführung in die Numerische Mathematik* Vol. 2 (Springer, Berlin, 1978).
- Schuster, H. G. *Deterministic Chaos* (Physik-Verlag, Weinheim, 1984).
- Delone, N. B., Zon, B. A. & Krainov, V. P. *Zh. eksp. teor. Fiz.* **75**, 445-453 (1978); *Soviet Phys. JETP* **48**, 223-227 (1978).
- Meerson, B. I., Oks, E. A. & Sasarov, P. V. *Pis'ma Zh. eksp. teor. Fiz.* **29**, 79-82 (1979); *JETP Lett.* **29**, 72-75 (1979).
- Delone, N. B., Krainov, B. P. & Shepelyansky, D. L. *Usp. Fiz. Nauk.* **140**, 355-392 (1983); *Soviet Phys. Usp.* **26**, 551-572 (1983).

Transcriptional autoregulation of the proto-oncogene *fos*

Paolo Sassone-Corsi, John C. Sisson & Inder M. Verma

Molecular Biology and Virology Laboratory, The Salk Institute, PO Box 85800, San Diego, California 92138, USA

Serum-induced transcription of the proto-oncogene fos is under negative feedback regulation mediated by the fos protein. The fos promoter region responsive to repression is also required for serum inducibility and binds a nucleoprotein complex in which the nuclear factor AP-1 is associated with fos protein.

TRANSCRIPTIONAL regulation of eukaryotic promoters is mediated by the interaction of specific *trans*-acting factors with various *cis*-regulatory elements¹⁻³. Several studies have demonstrated the importance of both the DNA-protein contacts and protein-protein interactions in obtaining efficient and specific initiation of transcription. The molecular mechanism by which the transcription machinery operates to generate tissue-specific, hormone-induced or growth/differentiation-related gene expression is still unclear. Characterization of the protein components involved in the formation of a transcriptional complex is an important step towards the understanding of gene regulation. Recently several transcriptional factors have been identified and characterized⁴. Interestingly, the phorbol ester induced factor AP-1 and the thyroid hormone receptor are the products of nuclear oncogenes *c-jun* and *erb-A*⁵⁻⁸. Another nuclear oncoprotein, *fos*, has been implicated as playing a pivotal role during cell growth, differentiation and development⁹⁻¹⁵. It was first identified as the cellular cognate of the oncogene *v-fos*, the

resident transforming gene of FBJ-murine osteosarcoma virus (FBJ-MSV)¹⁶. Both the viral and the cellular *fos* protein, despite different C-termini, can induce bone tumours *in vivo* and transform fibroblasts *in vitro*^{15,16}.

Recent results indicate that the *fos* protein may be involved in the regulation of the transcription of some genes^{12,17-19}. There is some evidence of *trans*-activation of mouse $\alpha 1$ collagen gene by the transiently expressed *v-fos* protein product²⁰. Activation of transcription of *Lex-A* operators has also been observed in yeast using bacterial *Lex-A*/mouse *fos* fusion proteins²¹. Additionally, *fos* or an antigenically related protein, has been found to be associated with the transcriptional complex of the adipocyte *aP2* gene promoter on which it apparently exerts a negative regulatory effect¹². Finally, *fos* protein contains sequences that might be involved in the recognition of specific promoter elements, as it shares limited sequence identities with the DNA binding domains of known transcriptional activators such as GCN4 and *v-jun*²².

Fig. 1 Repression of the serum-induced *c-fos* promoter activity by the *c-fos* protein. **a**, Structure of the human *c-fos* gene⁵³. The transcription initiation site (cap), the TATA box (▼), the exons (hatched boxes) and the position of several restriction sites in the upstream sequences are shown. The *NcoI* site in the fourth exon was used to create the frame-shift mutant VMM-FS²⁶ (see Fig. 2). FC3, FC4, FC5 and FC8 (ref. 54) are *fos*-CAT fusions in which *c-fos* promoter deletions from positions -711, -404, -307 and -220 to the +42 (site *NaeI*) were linked to CAT sequences. **b**, Expression of FC4 in serum starved (lane 1) and stimulated (lane 3) NIH 3T3 fibroblasts, after co-transfection, without (-) and with (+) the *c-fos*-expressing vector VMM²⁶ (see Fig. 2). **c**, Effect of cell confluency on repression by *c-fos* protein. In all cases the cells were serum-induced. The cell number per 10 cm plate at the time of DNA transfection is indicated; (-) and (+) as in **b**. RSV-CAT⁵⁶ was used as a transfection control. **Methods**. Transfection of plasmid DNA into NIH 3T3 mouse fibroblasts was performed using the calcium phosphate co-precipitation technique³². Cells were passaged in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS). Cells were transfected at the indicated density and exposed to the precipitate for 12-16 h. After washing, the medium was replaced with DMEM/0.5% FCS. After a further 48 h (at this stage we consider the cells to be serum-starved), cells to be stimulated were exposed to DMEM/20% FCS for a period of 60-180 min. CAT activity assays were performed as described⁵⁵. Quantitation of CAT activity was performed by densitometer scanning of several autoradiograms. In many experiments, a plasmid containing RSV LTR linked to the bacterial luciferase gene was co-transfected to monitor the transfection efficiencies.

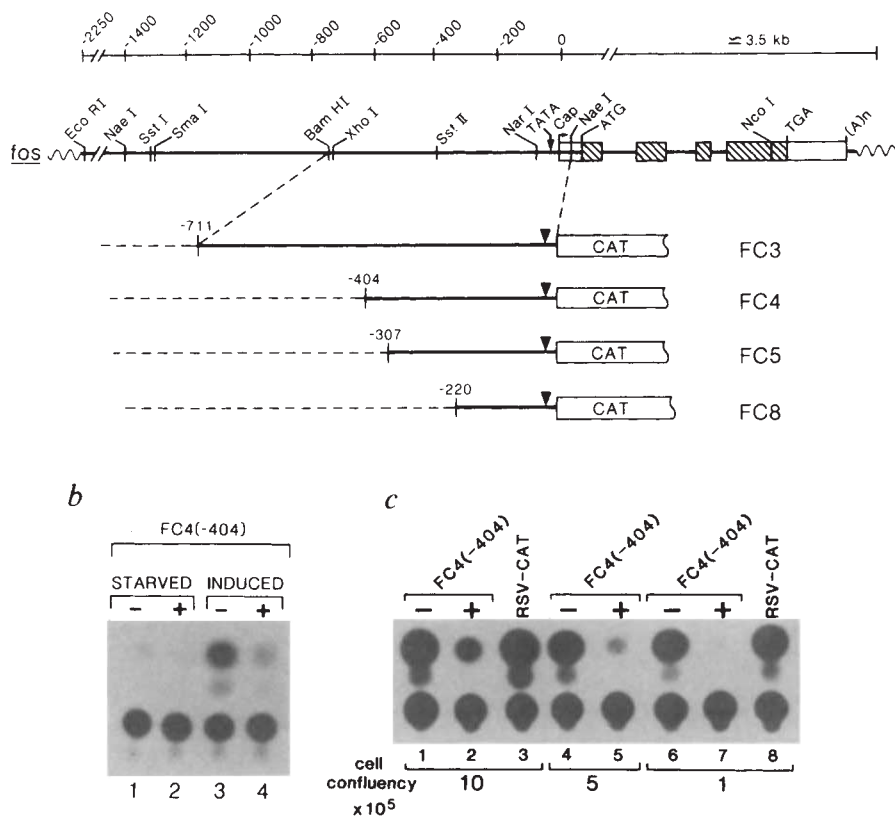
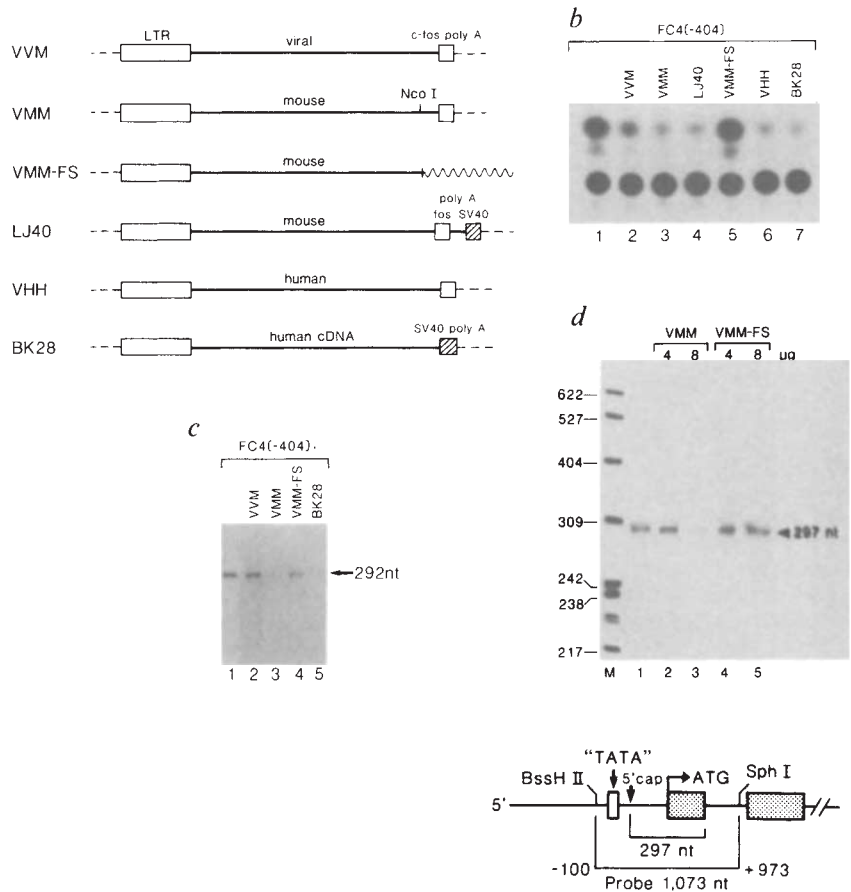


Fig. 2 Several *fos* expression vectors exert repression function. **a**, Structure of several *fos* chimaeric expression vectors (described in ref. 26) used in this study. VVM expresses a viral *fos* protein but contains the mouse *c-fos* 3' untranslated region. VMM expresses a mouse *c-fos* protein (p55) and VMM-FS is a VMM derivative in which the *Nco*I site located upstream from the carboxy domain has been filled with DNA-polymerase, creating a frame-shift mutation. LJ40 (ref. 56) is equivalent to VMM but contains a second poly(A) additional signal from SV40. VHH expresses a viral-human *fos* hybrid protein, whereas BK28 (ref. 57) contains the human *c-fos* full-length cDNA. All these coding sequences are under the transcriptional control of the FBJ-MSV provirus LTR²⁶. When these recombinants were transfected into NIH 3T3 fibroblasts, *fos* protein was detectable in similar quantities by immunoprecipitation with anti-*fos* monoclonal antibodies⁵⁸ and was localized to the nucleus by immunofluorescence (not shown). **b**, Co-transfection of FC4 with the *fos* expression vectors described in (a). The *c-fos* promoter is repressed by a factor of 8–10 for VMM, LJ40, VHH and BK28. The product of VVM represses with a much lower efficiency (about fourfold in this experiment), but in most cases no repression was observed), and VMM-FS shows no repression function. **c**, RNase protection mapping of RNA from cells transfected as in **b**. The results were consistent with those from the CAT assays were obtained; VMM and BK28 repress *c-fos* promoter activity, whereas both VMM-FS and VVM show no or little repression function. **d**, Repression of endogenous *c-fos* gene transcription by the *fos* protein product. The level of *c-fos* RNA in 60-min serum-induced HeLa cells is shown in lane 1. HeLa cells were transfected with 4–8 μ g of either VVM (lanes 2–3) or VMM-FS (lanes 4,5) and *c-fos* RNA levels were analysed.

Methods. Total cytoplasmic RNA was isolated and RNase mapping performed as described⁴⁴. Briefly, 30 μ g of RNA were hybridized with ³²P-labelled SP6 cRNA for 16 h at 45 °C in 80% formamide, 40 mM PIPES (pH 6.4), 400 mM NaCl and 1 mM EDTA. Following hybridization, samples were treated with a mixture of RNase A (40 μ g ml⁻¹) and RNase T1 (700 U ml⁻¹) at 32 °C for 60 min. The samples were further incubated for 15 min at 37 °C after addition of 1% SDS and 50 μ g of proteinase K, extracted with phenolchloroform, ethanol precipitated, and loaded on an 8% sequencing gel. The cRNA *fos*-CAT probe is 392 nucleotides(nt) long, from the *Eco*RI site in the CAT gene (+250) to the *Bss* HII site in the *c-fos* promoter (–100) and the expected size of the protected fragment is 292 nt, as 42 bp of the *c-fos* promoter are present in the FC4 fusion (see Fig. 1). For the endogenous RNA protection analysis, the cRNA probe is 1,073 nt (from the *Bss* HII site at –100 to the *Sph*I site at +973) and the expected size of the protected fragment is 297 nt (see the diagram in the lower panel). M is the size marker, pBR322/*Msp*I-digested and end-labelled. The intensity of the bands was measured by densitometric scanning of the autoradiogram.



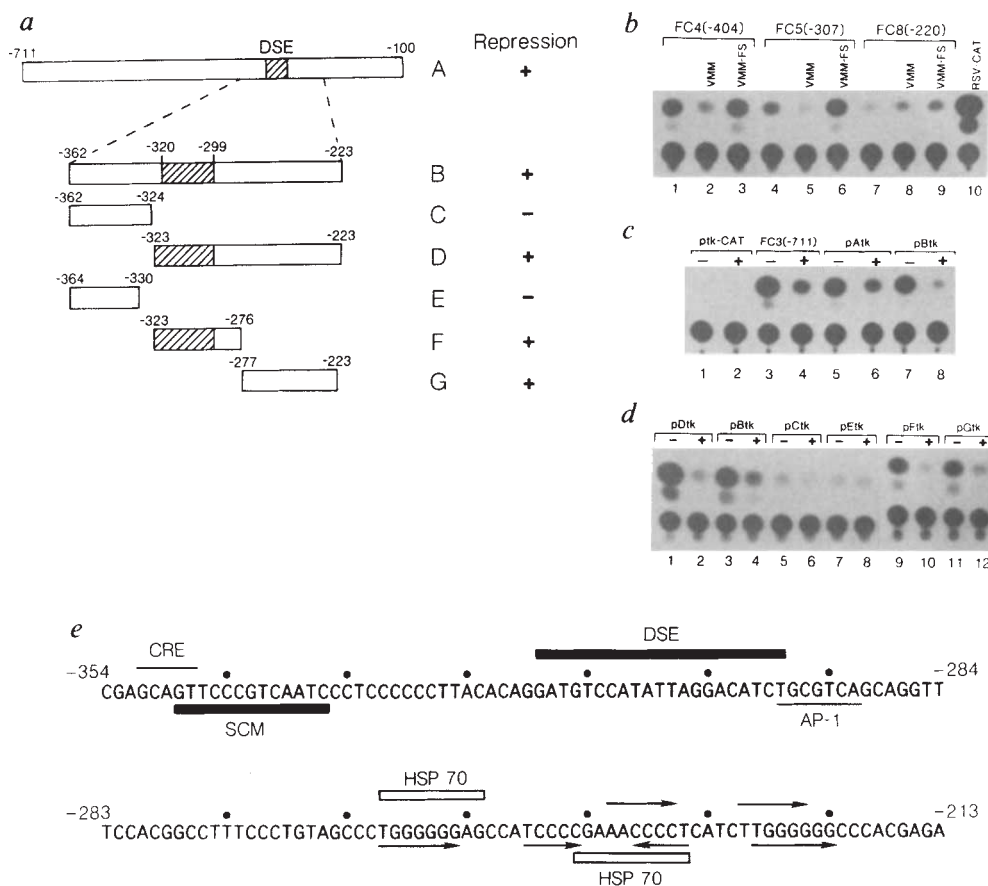
Transcription of *c-fos* gene is induced very rapidly by a large variety of mitogenic and differentiation-specific agents¹⁶. After induction, which does not require protein synthesis, transcription decreases rapidly, as shown by run-on experiments^{23–25}. Cycloheximide treatment delays the repression of *c-fos* transcription, indicating that a new serum-induced protein is required for the shut-off of *fos* transcription. We have investigated whether the *fos* protein itself could be the newly-synthesized protein that is responsible for the transcriptional repression of the *c-fos* promoter observed after serum induction. We report that the *c-fos* protein product specifically represses the serum-induced promoter activity of the *c-fos* gene. Such a feedback mechanism (*trans*-repression) occurs only when *c-fos* transcription is induced by serum and consequently only *c-fos* serum-inducible promoter elements are the molecular targets of the repression. *In vitro* binding studies indicate that the same serum-responsive promoter element that is repressed by the *c-fos* protein *in vivo* is associated with a nucleoprotein complex that contains the product of *c-fos* and the product of *c-jun*, the transcriptional factor AP-1.

Repression of promoter activity

To study the possible role of the *c-fos* protein in regulating the transcription of the *c-fos* promoter, we performed transient transfection assays. Plasmids containing upstream sequences of the *c-fos* gene linked to the bacterial chloramphenicol

acetyltransferase (CAT) gene (Fig. 1a) were co-transfected into NIH 3T3 fibroblasts with *c-fos* expression vectors. Figure 1b shows that construct FC4, containing 404 base pairs (bp) upstream of the *c-fos* start of transcription (see Fig. 1a), responds to induction with serum as judged by increased CAT-activity (compare lanes 1 and 3). This induction is considerably diminished or completely abolished if FC4 is co-transfected with plasmid VMM, which expresses a viral-mouse hybrid-*fos* protein²⁶ (compare lanes 3 and 4). Such repression is specific to the *c-fos* serum-induced promoter, as no repression was observed when the cells were serum starved (compare lanes 1 and 2). Diminution of *c-fos* promoter activity by *fos* protein is observed at several cell confluencies, although the most dramatic effect is observed after transfection of cells seeded at lower density (Fig. 1c). The serum-induced level of FC4 as well as the basal transcription of the RSV-CAT control are very similar at different cell densities. Figure 2a shows the structure of a number of different recombinants expressing mouse or human *c-fos* protein products. Figure 2b (lane 2) shows that viral *fos* protein (VVM) is also able to repress the *c-fos* promoter, albeit at a lower efficiency. In several other experiments the level of repression by *v-fos* is negligible (Fig. 2c, lane 2 and data not shown). Repression of the *c-fos* promoter activity is a property of the *c-fos* protein and not dependent on the nature of the sequences in the non-coding domain because both VMM (lane 3) and LJ40 (lane 4), which differ in their 3' non-coding domain, have similar

Fig. 3 Identification of *c-fos* promoter sequences involved in the trans-repression. *a*, Fragments of the upstream region of the *c-fos* promoter were purified and cloned upstream from the promoter in a herpes thymidine kinase/CAT recombinant (equivalent to TK-CAT in Fig. 4). The fragments are represented in the figure and their termini are indicated with respect to the *c-fos* cap-site (+1). The hatched region represents the dyad symmetry element (DSE). *b*, Co-transfection experiments using several *c-fos* promoter deletion mutants (see Fig. 1*a*) together with *fos* expression vectors (see Fig. 2*a*). RSV-CAT is a transfection control. *c* and *d*, Transfection experiments without (–) and with (+) co-transfection of VMM (Fig. 2*a*). *c*, ptk-CAT is equivalent to TK-CAT in Fig. 4 and contains no *c-fos* promoter sequences. FC3 is described in Fig. 1*a*. Recombinants pAtK and pBtK contain the A and B fragments, respectively, linked upstream from the TK/CAT fusion gene. *d*, Recombinants pCtK, pDtK, pEtK, pFtK and pGtK contain, respectively, fragments C, D, E, F and G linked upstream from the TK/CAT fusion gene. *e*, Sequence of fragment B, between positions –362 and –223. Several promoter landmarks are indicated: DSE (dyad symmetry element³³), SCM (sis-conditioned medium responsive element³¹), CRE (a cyclic-AMP-like responsive element), AP-1 (an AP-1-like binding site). The region between –277 and –223 (corresponding to fragment G in Fig. 3*a*) is rich in sequence symmetries and repeats, which are indicated by the arrows. The regions of homology to the human HSP70 promoter are also indicated.



effects. As VMM, LJ40 and VHH all generate viral and cellular hybrid *fos* proteins, we tested a human *c-fos* cDNA clone (BK28, lane 7) which also yielded similar results. A frame-shift mutant (VMM-FS), at the carboxyl end of the *fos* protein does not repress *c-fos* promoter activity (lane 5), indicating that trans-repression by *fos* protein may require an authentic C-terminus (VMM-FS does, however, synthesize a nuclear protein as judged by immunofluorescence and immunoprecipitation with anti-*fos* antibodies, unpublished data). Figure 2*c* shows that repression by the *c-fos* protein product is also detected by analysing the *fos*-CAT RNA. RNA from serum-induced cells co-transfected with FC4 and various *fos* expression vectors was analysed by RNase protection analysis using a *fos*-CAT SP6 cRNA probe. The results are equivalent to the data in Fig. 2*b*.

Repression of endogenous transcription

The *c-fos* product represses *c-fos* promoter activity when the promoter is introduced into culture cells as part of an episomal construct. We next tested whether *c-fos* protein also represses transcription of the serum-induced endogenous *c-fos* gene. HeLa cells were transfected with VMM (Fig. 2) and total RNA was isolated and analysed by RNase protection analysis. Data in Fig. 2*d* show that when 8 μ g of VMM is transfected into cells, the level of *c-fos* RNA detected after serum induction (lane 3) is only 15% of that of the control (lane 1). Furthermore, the VMM frame-shift mutant has no effect on endogenous *c-fos* transcription (lanes 4 and 5). We therefore conclude that the *fos* protein product can trans-repress transcription of its own promoter when this is situated in its normal genomic environment.

Delineation of responsive promoter sequences

As repression of the *c-fos* promoter is observed only after serum induction, we expected the dyad symmetry element (DSE), which is required for serum, TPA and EGF induction^{27–30} to be one of the molecular targets involved. We therefore analysed a number of recombinants, containing different portions of the *c-fos* upstream region, to test this hypothesis. Deletion analysis of the *fos*-CAT fusion plasmid showed that the region between positions –404 (FC4) and –220 (FC8), is the target for repression by the *fos* protein (Fig. 3*b* and *c*, lanes 3, 4). This region contains several promoter elements required for basal and induced expression (Fig. 3*e*): the DSE, an AP-1-like site, the *sis* conditioned medium responsive element³¹, a cyclic-AMP responsive element consensus, the adenovirus E1A-responsive region³² and the HTLV-I tat-I responsive region (M. Fujii, personal communication). To examine which transcriptional element(s) is responsible for the repression effect, we cloned several subfragments (see Fig. 3*a*) of the *c-fos* promoter upstream from a herpes thymidine kinase/CAT gene recombinant (equivalent to the vector described in Fig. 4*a*). A TK/CAT recombinant containing a large fragment (–711 to –100, fragment A in Fig. 3*a*) of the *c-fos* promoter shows that the *c-fos* protein product exerts its repression effect even in a heterologous promoter environment. Serum-induced transcription from a recombinant containing fragment B (–362 to –223; Fig. 3*d*) is also repressed by *c-fos* protein. Serum-inducibility is conserved by fragment D (Fig. 3*a*) as well as by its derivatives, fragments F and G. That F is serum-inducible is not surprising, as it contains the DSE^{27,28,33}, whereas element G has not previously been shown to exhibit characteristics of serum-induci-

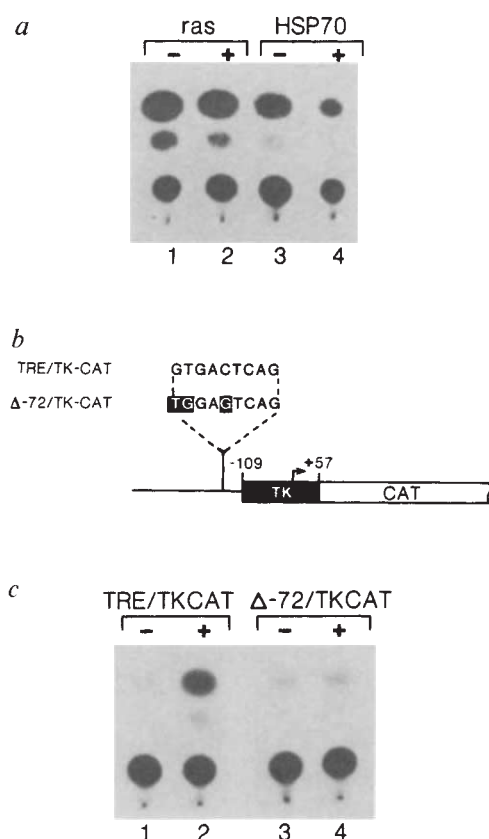


Fig. 4 Repression of HSP70 promoter and activation of a TPA responsive element by co-transfection with *c-fos* expression vector. *a*, Co-transfection of the *c-Harvey-ras* promoter/CAT recombinant with (+) or without (–) VMM showed no alteration in promoter activity. The *c-Ha-ras* promoter region used covers positions –420 to +130. The activity of the HSP70 promoter⁵⁹ was decreased on co-transfection with VMM. The HSP70 promoter region used contains sequences between positions –1,100 to +150. *b*, Structure of the recombinants used. The synthetic oligodeoxynucleotide bearing the human metallothionein IIa gene TRE sequence was cloned upstream from the TK-CAT transcription unit (TRE/TK-CAT). A three-point mutation of this TRE was also generated and cloned in the same position³⁵ (Δ -72/TK-CAT). *c*, Transfection experiments in NIH 3T3 fibroblasts with (+) or without (–) VMM (see Fig. 2a) and either TRE/TK-CAT or Δ -72/TK-CAT. No treatment with either serum or TPA was performed on the transfected cells. Serum treatment shows no effect, whereas TPA treatment slightly increases the activation by *c-fos* product (data not shown).

bility. From our results this fragment, linked to the TK/CAT fusion gene, confers serum-inducibility to the TK promoter with an efficiency comparable to fragment F, which contains the DSE. Interestingly, fragment G contains several sequence repeats which may be required for serum-inducibility (see sequence, Fig. 3e). The region upstream from position –323 is not serum-responsive (fragments C and E) and is not trans-repressed by *fos* products. As shown in Fig. 3c and 3d, only the serum-responsive *c-fos* elements are targets for repression by *c-fos* products.

Repression of HSP70 promoter activity

To evaluate the generality of the action of the *fos* protein in regulating transcription, we studied a number of promoters. Most promoter activities were unaffected by co-transfection of a *c-fos* expression vector in NIH 3T3 cells; among the different promoters analysed were SV40 early, adenovirus E2a and E3, RSV LTR, adenosine deaminase, human *c-myc*, stromolysin, HTLV-I LTR and *c-Harvey-ras* (data for *c-Ha-ras* is shown,

Fig. 4a, lanes 1, 2). On the other hand, transcription from the human heat-shock promoter³⁴ (HSP70), which is also serum-responsive³⁴, is repressed by the *c-fos* protein product (Fig. 4a, compare lanes 3 and 4). A search for sequence similarities between *c-fos* and HSP70 promoters revealed some homology between a region of fragment G and an HSP70 element located between positions –110 and –120 and positions –132 and –140 (see Fig. 3e).

Induction of TPA-responsive element

The data shown above indicate that *c-fos* promoter activity is repressed by *c-fos* protein. As it has been reported that *fos* protein is associated with the nuclear transcription factor AP-1^{12,17,18}, we undertook experiments to determine whether *fos* protein influences the transcription of a characterized TPA-responsive element (TRE, binding site AP-1^{12,17,18}). Note the AP-1-like sequence in the *fos* promoter. Figure 4c show that when a recombinant containing the metallothionein II TPA responsive element linked to a TK/CAT fusion gene (TRE/TK-CAT, see Fig. 4b) is co-transfected with a *c-fos* expression vector, the CAT activity increases by a factor of 9–10 (lanes 1, 2). No induction (Fig. 4c, lanes 3, 4) is observed when a mutant TRE is used (Δ -72/TK-CAT, Fig. 4b). The same mutation in the AP-1 site also impairs TPA responsiveness³⁵ (TPA treatment only slightly increases the induction by the *c-fos* protein.) Thus in this case, the *c-fos* protein activates the transcription of a TPA-responsive element in the absence of TPA.

The *c-fos* promoter transcriptional complex

Our results indicate that the *c-fos* protein product may be directly involved in regulating transcription from its own promoter as well as mimicking the effects of TPA induction. Recent results from other laboratories indicate that the *fos* protein appears to be associated with other nuclear proteins in transcriptional complexes^{12,17–19}. We therefore performed *in vitro* binding studies by gel retardation analysis using nuclear extracts from HeLa cells and an end-labelled probe equivalent to fragment D (see Fig. 3a). As shown in Fig. 5 (lanes 1, 10), several nucleoprotein complexes are formed with fragment D; the two major complexes are indicated as A and B. Complex A was previously observed by others and was demonstrated to contain the serum-responsive factor (SRF)^{27,29,36}. Preincubation of the nuclear extract with anti-*c-fos* polyclonal antibodies (M2Ab) raised against a *fos* peptide³⁷ (M2) decreases the formation of complex A (lane 2), indicating that the *c-fos* protein product directly interacts with the transcriptional complex that dictates the serum response. When a fivefold excess of M2 peptide was added together with the M2Ab in the binding reaction, no disruption of complex A was observed (lane 3). Similar experiments were performed using *c-fos* monoclonal antibodies raised against the N-peptide (amino acids 4–17), and similar results were obtained (lane 4). Furthermore, competition with the N-peptide abolished the disruption of complex A (lane 5). That nucleoprotein complex A also contains the transcriptional factor AP-1 is demonstrated by pre-treating nuclear extracts with anti-PEP-1 and anti-PEP-2 antibodies raised against the *jun* protein, which have been shown to selectively immunoprecipitate the AP-1 protein⁵ (lanes 7, 9). Extracts treated with pre-immune antisera for PEP-1 and PEP-2 show the same pattern of retarded bands as the control (lanes 6, 8). Thus it appears that formation of complex A must involve both the nuclear transcription factor AP-1 and *fos*. In control experiments, we also used antibodies against other nuclear proteins, some of which are known to be transcriptional factors, such as the glucocorticoid receptor³⁸ and the thyroid receptor^{7,8} or p53 (ref. 39). These antibodies did not interfere with the formation of complex A (lanes 12–14). These results indicate that *fos* protein directly participates in the formation of a transcriptional complex on its own promoter. The promoter region involved is the same that has been shown to be the target of *in vivo* repression by the *fos* product.

Discussion

The *c-fos* gene is a prototype of the immediate early genes that are activated in response to a diverse array of agents including mitogens, growth factors, differentiation inducers and biomodulators.⁴⁰ Upon induction, the transcription rate of these early response genes increases very rapidly and then, in most cases, decreases precipitously.²³ The products of a large majority of these genes are nuclear, suggesting a possible role in the regulation of gene expression. Not surprisingly some of them have been demonstrated to be transcriptional factors⁵⁻⁸. The transcription profile of many of these genes is dramatically altered in the presence of inhibitors of protein synthesis^{23,41-43}. In the case of *fos*, for instance, cycloheximide or anisomycin treatment leads to superinduction of *fos* mRNA transcripts which suggest that after mitogen treatment, *de novo* protein synthesis is not required for *fos* expression, and that pre-existing nuclear factors activate *fos* transcription^{30,40}. More importantly, in the presence of inhibitors of protein synthesis, transcription from the *fos* promoter is prolonged (4–6 h compared to 30–60 min). It thus appears that repression of *c-fos* gene transcription has an absolute requirement for *de novo* synthesized protein(s), which could either directly modulate *c-fos* transcription or, alternatively, modify other nuclear proteins contributing toward the regulation of *c-fos* gene expression. We show that the *c-fos* product itself fulfills the properties of the above postulated *de novo* synthesized protein. In co-transfection experiments, the *c-fos* product represses serum-induced transcription from the *c-fos* promoter (Figs 1 and 2). The repressor function of the *c-fos* protein is not limited to the episomal state of the transfected DNA, but is also seen when the endogenous *c-fos* gene is in its natural genomic environment (Fig. 2). Finally, repression is also observed when the *c-fos* serum-responsive promoter region is linked in either orientation to a heterologous promoter, which by itself is otherwise unaffected by co-transfection with a *c-fos* expression vector.

Two adjacent regions of the *c-fos* promoter have been identified to be responsive to repression; the first (contained in fragment F in Fig. 3a) contains the DSE and an AP-1-like site. The second (contained in fragment G in Fig. 3a) contains several sequence repeats and dyad symmetries (see Fig. 3e) and displays sequence identities with the human HSP70 promoter. Interestingly, the HSP70 gene is serum responsive³⁴ as well as being a target of *trans*-repression by the *c-fos* protein product (Fig. 4a). These cumulative observations lend credence to a mechanism of feedback regulation in which transcription of *c-fos* is repressed by its own product. Such a mechanism would account for the dramatic decrease in the *c-fos* transcriptional rate after the peak of induction. As most of the genes that belong to the 'serum early response' class are regulated with similar transcriptional kinetics⁴⁵, it is possible that the *c-fos* protein product negatively regulates their transcription, or that their products have transcriptional characteristics similar to *fos* protein. Some of the serum-inducible early-response genes have been shown to code for positive trans-regulators related to the AP-1 protein⁴⁶ (L. Lau and R. Bravo, personal communication). These observations indicate that an early response to growth signals involves the induction of a complex array of transcriptional inducers and repressors which are likely to 'cross-talk' in a complicated regulatory network.

Association of *fos* and *jun*

What is the molecular mechanism by which *fos* regulates transcription? Although the *fos* protein is associated with DNA^{47,48} and has some sequence identity with DNA-binding domains of nuclear factors such as AP-1 and GCN4 (ref. 22), we postulate that it acts by interacting with DNA-binding transcriptional factors. Recent observations support this hypothesis, as *fos*-associated cellular protein p39 is structurally related to AP-1 (refs 18, 19, 49). Furthermore, no direct binding of purified p55

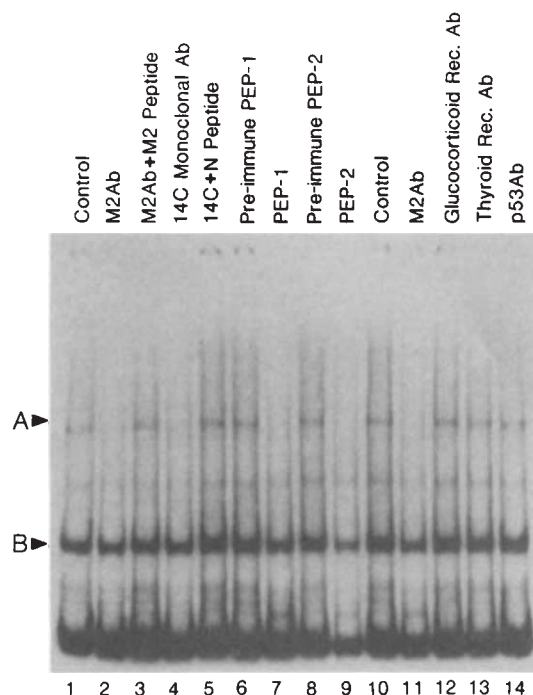


Fig. 5 Both *c-fos* and *c-jun* products are present in the *c-fos* promoter transcriptional complex. Gel shift analysis of nucleoprotein complexes formed on the *c-fos* promoter fragment D (see Fig. 2a) using nuclear extract from HeLa cells. Both polyclonal (lane 2) and monoclonal (lane 4) anti-*c-fos* antibodies prevent complex A formation. Inhibition of DNA-protein complex binding with antibodies was performed by adding an excess of either *c-fos* M or N-peptide to the preincubation. *Jun* antibodies PEP-1 and PEP-2 (ref. 5) also prevent the formation of complex A (lanes 7, 9), whereas control antibodies against the glucocorticoid receptor, the thyroid receptor and p53 (lanes 12–14) do not prevent the formation of complex A. Pre-immune sera for PEP-1 and PEP-2 (lanes 6, 8) show the same binding profile as the untreated extract.

Methods. HeLa cell nuclear extracts were made essentially as described⁶⁰. Final protein concentrations were typically 1–2 mg ml⁻¹. DNA-protein binding was conducted in 20 µl incubation. Nuclear extract (1–3 µl) was incubated with 1 µg of poly (dI:dC) in binding buffer (10 mM Tris-HCl [pH 7.5], 50 mM NaCl, 5% glycerol, 5 mM MgCl₂, 1.0 mM DTT and 1.0 mM EDTA) for 20 min at room temperature. When required, nuclear extracts were incubated overnight at 4°C with different antibodies. The pre-incubations with *c-fos* peptide antibodies with M- or N-peptides were conducted for 60 min at 4°C (about 0.5 µg of antibody to 5 µg of peptide). The complexes were resolved on non-denaturing 4% polyacrylamide gels.

to the *fos* promoter has been observed¹⁸. Data presented here demonstrate that the *fos* protein is required for assembly of the nucleoprotein complex associated with the *c-fos* serum-responsive promoter region, and that *fos* is therefore likely to interact with AP-1 and the serum responsive factor (SRF). The molecular mechanism by which the *fos* protein represses transcription remains obscure, but it is possible that the *fos* protein acts by displacing SRF, which is required for efficient induced expression^{30,33}, and/or that the *fos* protein modifies, either directly or indirectly, the transcriptional factors required for full expression. The *fos* protein has also been found in the transcriptional complex involved in the regulation of the adipocyte gene, aP2 (ref. 12) which contains a TRE. Although it is unclear what physiological function *fos* protein could have in this context, it appears to repress transcription of the aP2 gene¹².

Promoter elements lacking AP-1 binding sites can also be regulated by *fos* protein. For instance, the *fos* serum-responsive promoter fragment G (see Fig. 3), linked to the TK/CAT fusion

gene, is strongly repressed by the fos products, although the sequence of the G fragment (see Fig. 3e) shows no homology to an AP-1 binding site.

Pleiotropic functions of fos protein

The fos protein is clearly involved in the repression of its own transcription and that of the $\alpha 2$ promoter^{12,17}. Data from experiments using an isolated TPA-responsive element (Fig. 4a), however, show that fos protein can also exert a positive trans-regulatory function. Although fos protein appears to be consistently associated with AP-1 (see discussion above), it is conceivable that the activity of the fos/AP-1 complex in regulating transcription is modulated by yet other nuclear factors. The manner in which fos modulates AP-1 activity remains unclear. It has been previously reported that, upon TPA treatment, both fos and the associated polypeptide p39 (AP-1) are modified^{50,51}. The covalent modification of p39, possibly by phosphorylation, could be mediated by the fos protein. As the AP-1 binding site is symmetric, the AP-1 protein could bind to TRE as a dimer and fos might catalyse the dimer formation and/or stabilize the binding.

To examine the structural and functional relationships between fos protein and AP-1, we have generated fos mutants. Initial results indicate that intact amino and carboxy termini of the fos protein are crucial for its repression function (unpublished data). It appears that more extensive changes are required in the carboxy-terminal domain to alter fos function because linker insertions in the C-terminal region only slightly alter the trans-regulatory activity (data not shown). The carboxy-terminal domain of the fos protein contains the major sites for the covalent modifications observed upon TPA treatment⁵¹. The v-fos protein product has an altered carboxy terminus and has

previously been shown to act as a trans-activator of the mouse α -1 collagen promoter²⁰. Using a large number of promoters, including c-fos (see Fig. 2), we have not been able to detect a reproducible trans-regulatory effect of the v-fos protein. As v-fos protein also binds to p39 (AP-1) and can transform cells^{50,51}, we presume that the carboxy-terminal domain of the fos protein is not essential for either interaction with p39 or cellular transformation. The separation of the trans-regulatory and transformation properties of the fos protein is reminiscent of that previously reported for adenovirus 2 E1A protein⁵².

Regulation of expression of proto-oncogenes is central to understanding their role in normal cells and oncogenesis. In response to growth stimuli, the cell reacts by initiating a complex transcriptional program. Some of the nuclear trans-acting factors involved in transcription have been shown to be previously-identified oncogenes. Other nuclear proteins may not directly interact with target DNA sequences, but could influence expression by protein-protein interactions. The ability of the fos protein to autoregulate its transcription may serve as a paradigm to study regulation of other nuclear oncogenes.

We thank T. Bos and P. Vogt for the gift of PEP-1 and PEP-2 antibodies, P. Angel and M. Karin for plasmids TRE/TK CAT and Δ -72-TK CAT and Drs W. W. Lamph, M. Fujii, C. Thompson, S. Hollenberg, L. Tack, J. Visvader, G. Merlino, W. Morgan, M. Vogt and V. Raymond for other reagents. We thank Drs P. De Togni and J. Barber for discussions and W. W. Lamph and M. McKeown for reading the manuscript. P.S.-C. is partially supported by the CNRS and is on leave of absence from the Laboratoire de Genetique Moleculaire des Eucaryotes, CNRS, Strasbourg, France. This work was supported by grants from the NIH and American Cancer Society, Inc. to I.M.V.

Received 28 April; accepted 10 June 1988.

- Dynan, W. S. & Tjian, R. *Nature* **316**, 774-778 (1985).
- Sassone-Corsi, P. & Borrelli, E. *Trends Genet.* **2**, 215-219 (1986).
- Jones, N. C., Rigby, P. W. J. & Ziff, E. B. *Genes Dev.* **2**, 267-281 (1988).
- Wingender, E. *Nucleic Acids Res.* **16**, 1879-1902 (1988).
- Bohmann, D. et al. *Science* **238**, 1386-1392 (1987).
- Angel, P. et al. *Nature* **332**, 166-171 (1988).
- Sap, J. et al. *Nature* **324**, 635-640 (1986).
- Weinberger, C. et al. *Nature* **324**, 641-646 (1986).
- Holt, J. T., Venkat Gopal, T., Moulton, A. D. & Nienhuis, A. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4794-4798 (1986).
- Nishikura, K. & Murray, J. M. *Molec. cell. Biol.* **7**, 639-649 (1987).
- Muller, R. & Wagner, E. *Nature* **311**, 438-442 (1984).
- Distel, R. J., Ro, H.-S., Rosen, B. S., Groves, D. L. & Spiegelman, B. M. *Cell* **49**, 835-844 (1987).
- Ruther, U., Garber, C., Komitowski, D., Muller, R. & Wagner, E. F. *Nature* **325**, 412-416 (1987).
- Dony, C. & Gruss, P. *Nature* **328**, 711-714 (1987).
- Verma, I. M. & Graham, W. R. *Adv. Cancer Res.* **49**, 29-52 (1987).
- Verma, I. M. *Trends Genet.* **2**, 93-96 (1986).
- Rauscher, F. J. III, Sambucetti, L. C., Curran, T., Distel, E. J. & Spiegelman, B. M. *Cell* **52**, 471-480 (1988).
- Sassone-Corsi, P., Lamph, W. W., Kamps, M. & Verma, I. M. *Cell* (in the press).
- Franza, R. B., Jr., Rauscher, F. J. III, Josephs, S. F. & Curran, T. *Science* **239**, 1150-1153 (1988).
- Sassone-Corsi, P., Frunzio, R., Lian, G., Mudryi, M. & De Crombrughe, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3213-3217 (1986).
- Lech, K., Anderson, K. & Brent, R. *Cell* **52**, 179-184 (1988).
- Vogt, P., Bos, T. J. & Doolittle, R. F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3316-3319 (1987).
- Greenberg, M. E. & Ziff, E. B. *Nature* **311**, 433-438 (1984).
- Greenberg, M. E., Hermanowski, A. L. & Ziff, E. B. *Molec. cell. Biol.* **6**, 1050-1057 (1986).
- Mitchell, R. L., Henning-Chubb, C., Huberman, E. & Verma, I. M. *Cell* **45**, 497-504 (1986).
- Miller, A. D., Curran, T. & Verma, I. M. *Cell* **36**, 51-60 (1984).
- Gilman, M., Wilson, R. N. & Weinberg, R. A. *Molec. cell. Biol.* **6**, 4305-4316 (1986).
- Greenberg, M. E., Siegfried, Z. & Ziff, E. B. *Molec. cell. Biol.* **7**, 1217-1225 (1987).

- Treisman, R. *EMBO J.* **6**, 2711-2717 (1987).
- Treisman, R. *Cell* **46**, 567-574 (1986).
- Hayes, T. E., Kitchen, A. M. & Cochran, B. H. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1272-1276 (1987).
- Sassone-Corsi, P. & Borrelli, E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6430-6433 (1987).
- Treisman, R. *Cell* **42**, 889-902 (1985).
- Wu, B. J. & Morimoto, R. I. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6070-6074 (1985).
- Angel, P. et al. *Cell* **49**, 729-739 (1987).
- Fisch, T. M., Prywes, R. & Roeder, R. G. *Molec. cell. Biol.* **7**, 3490-3502 (1987).
- Curran, T., Van Beveren, C., Ling, N. & Verma, I. M. *Molec. cell. Biol.* **5**, 107-112 (1985).
- Giguere, V., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. *Cell* **46**, 645-652 (1986).
- Gurney, E. G., Harrison, R. O. & Fenno, J. J. *Virology* **34**, 752-763 (1980).
- Verma, I. M. & Sassone-Corsi, P. *Cell* **51**, 513-514 (1987).
- Cochran, B. H., Reffel, A. C. & Stiles, C. D. *Cell* **33**, 939-947 (1983).
- Kruijer, W., Cooper, J. A., Hunter, T. & Verma, I. M. *Nature* **312**, 711-716 (1984).
- Muller, R., Bravo, R., Burckhardt, J. & Curran, T. *Nature* **312**, 716-720 (1984).
- Sassone-Corsi, P. & Verma, I. M. *Nature* **326**, 507-510 (1987).
- Lau, L. F. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1182-1186 (1987).
- Ryder, K., Lau, L. F. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1487-1491 (1988).
- Sambucetti, L. & Curran, T. *Science* **234**, 1417-1419 (1986).
- Renz, M., Vermier, B., Jurz, C. & Muller, R. *Nucleic Acids Res.* **15**, 277-293 (1987).
- Rauscher, F. J. III et al. *Science* **240**, 1010-1016 (1988).
- Curran, T., Miller, A. D., Zokas, L. & Verma, I. M. *Cell* **36**, 259-268 (1984).
- Barber, J. & Verma, I. M. *Molec. cell. Biol.* **7**, 2201-2211 (1987).
- Lillie, J. W., Green, M. & Green, M. R. *Cell* **46**, 1043-1051 (1986).
- Van Straaten, F., Muller, R., Curran, T., Van Beveren, C. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3183-3187 (1983).
- Deschamps, J., Meijlink, F. & Verma, I. M. *Science* **230**, 1174-1177 (1985).
- Gorman, C., Moffat, L. F. & Howard B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
- Meijlink, F., Curran, T., Miller, A. D. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4987-4991 (1985).
- Verma, I. M. et al. *Gene Transfer in Animals* (UCLA Symposium, in the press).
- De Togni, P., Niman, H., Raymond, V., Sawchenko, P. & Verma, I. M. *Molec. cell. Biol.* (in the press).
- Ishii, S., Merlino, G. & Pastan, I. *Science* **230**, 1378-1381 (1985).
- Dignam, J. D., Lebovitz, R. M. & Roeder, R. *Nucleic Acids Res.* **11**, 1475-1489 (1983).

Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA

Jerry Pelletier* & Nahum Sonenberg*†

Department of Biochemistry* and McGill Cancer Center†, McGill University, Montreal, Quebec H3G 1Y6, Canada

Poliovirus RNA is naturally uncapped, therefore its translation must proceed via a cap-independent mechanism. Translation initiation on poliovirus RNA occurs by binding of ribosomes to an internal sequence within the 5' noncoding region. This novel mechanism of initiation may explain the disparate translation of several other eukaryotic messenger RNAs.

THE mechanism of initiation of translation on poliovirus RNA is enigmatic. Poliovirus RNA, unlike most eukaryotic mRNAs is not blocked at its 5' end by the cap structure, m⁷GpppX (where X is any nucleotide). Instead, a small polypeptide termed VPg is covalently linked to the 5' end of genomic RNA^{1,2} and is removed in the cytoplasm³. The 5' end of polysomal RNA terminates in pUp^{4,5}. Poliovirus RNA possesses an unusually long 5' untranslated region (UTR) (~750 nucleotides) which is highly conserved among the three poliovirus serotypes⁶⁻⁸. This region contains 7-8 AUGs, of which only three are conserved in position, although the open reading frames (ORFs) specified by them are not conserved in length or amino-acid content⁸. The translation products originating from these ORFs have not been observed *in vivo* or *in vitro* and are not necessary for viral biogenesis (J.P., M. E. Flynn, G. Kaplan, V. R. Racaniello and N.S., manuscript submitted).

A scanning model has been proposed to explain the mechanism by which eukaryotic mRNAs initiate translation⁹. According to this model, ribosomes and associated factors bind at or near the 5' end of the mRNA in a process that is facilitated by the presence of a cap structure, and scan the mRNA until the appropriate initiator codon is reached⁹. Two modifications were introduced to the model to explain translation on polycistronic eukaryotic mRNAs. It was postulated that ribosomes can terminate translation at an upstream ORF, then resume scanning and initiate translation at the downstream ORF¹⁰⁻¹² and that ribosomes could skip an upstream AUG and initiate at a downstream AUG (the latter process was termed leaky scanning¹²). In either case, the scanning model excludes the possibility of independent internal ribosome binding on eukaryotic mRNAs¹³.

The translation of poliovirus RNA is not satisfactorily accommodated by the scanning model. Because poliovirus RNA is naturally uncapped, it must translate by a cap-independent mechanism. We have used deletion mutagenesis to show that an internal sequence in the 5' UTR of poliovirus RNA is required for cap-independent translation¹⁴ and this sequence can also confer cap-independent translation to heterologous mRNAs¹⁴. This raises the interesting possibility that ribosomes can bind internally to this region of the poliovirus 5' UTR. But, in these experiments the possibility that ribosomes bind to the 5' end of poliovirus mRNA and then migrate towards the region important for cap-independent translation was not rigorously excluded.

We designed experiments to test the hypothesis that an internal sequence of the 5' UTR of poliovirus mRNA can mediate internal initiation. An assay was devised whereby the poliovirus 5' UTR could direct translation initiation independently of 5'-end-mediated initiation. The poliovirus 5' UTR was inserted as the intercistronic spacer in an artificially created bicistronic mRNA (Fig. 1a). The bicistronic plasmid contains as the first cistron the herpes simplex virus-1 thymidine kinase (TK) gene

and as the second cistron, the bacterial *chloramphenicol acetyltransferase* (CAT) gene (Fig. 1a). A control plasmid contained the 5' UTR of CAT as the intercistronic spacer. The rationale for the experiment is that if the poliovirus 5' UTR mediates internal initiation, then under conditions which abolish initiation at the first cistron (for example, in poliovirus-infected cells), translation of the second cistron (CAT) in TK/P2CAT should be unaffected (Fig. 1a).

Translation in poliovirus-infected cells

To determine whether internal initiation could occur *in vivo*, the bicistronic constructs TK/CAT and TK/P2CAT were subcloned into the pSV2 expression vector. The pSV2 constructs were transfected into COS-1 cells and TK and CAT protein levels were determined by pulse-labelling with [³⁵S]methionine, followed by immunoprecipitation analysis. Parallel dishes of transfected COS-1 cells were infected with poliovirus and labelled with [³⁵S]methionine four hours post-infection; at this time cellular mRNA translation is abrogated because of inactivation of eIF-4F (a translation initiation factor required for cap-dependent mRNA translation)¹⁵. Antisera directed against TK immunoprecipitated a polypeptide with a relative molecular mass (*M_r*) of ~46,000 (46K) which corresponds to HSV-1 TK in extracts from cells which had been transfected with pSV2/TK/CAT and pSV2/TK/P2CAT (Fig. 1b, lanes 2 and 3 respectively), but not from extracts of mock-transfected COS-1 cells (lane 1). The TK product was not detected in pSV/TK/CAT and pSV/TK/P2CAT-transfected COS-1 cells which were subsequently infected with poliovirus (Fig. 1b, lanes 4 and 5 respectively). Anti-CAT antibodies immunoprecipitated a polypeptide of ~25K from extracts of cells transfected with pSV/TK/CAT (lane 7) and pSV/TK/P2CAT (lane 8) but not from extracts of mock-transfected cells (lane 6). Following poliovirus infection, CAT synthesis was detected in pSV2/TK/P2CAT (lane 10), but not in pSV2/TK/CAT (lane 9) transfected COS-1 cells. Because poliovirus infection blocks 5'-end mediated initiation on the TK gene, CAT synthesis observed in poliovirus-infected COS-1 cells transfected with pSV/TK/P2CAT (lane 10) is not the result of termination/reinitiation events or leaky scanning, but is probably due to internal binding of ribosomes.

We used another strategy to substantiate the finding that translation initiation at the two cistrons occurs by two independent mechanisms. Translation of most cellular mRNAs is reduced when cells are incubated in hypertonic conditions. It has been previously shown that poliovirus translation is more resistant to hypertonic inhibition than cellular mRNAs^{16,17}. We reasoned that under hypertonic conditions, translation of the TK cistron should be diminished to a greater extent than the downstream CAT cistron which is preceded by the poliovirus 5' UTR. Figure 1c shows the synthesis of TK and CAT proteins

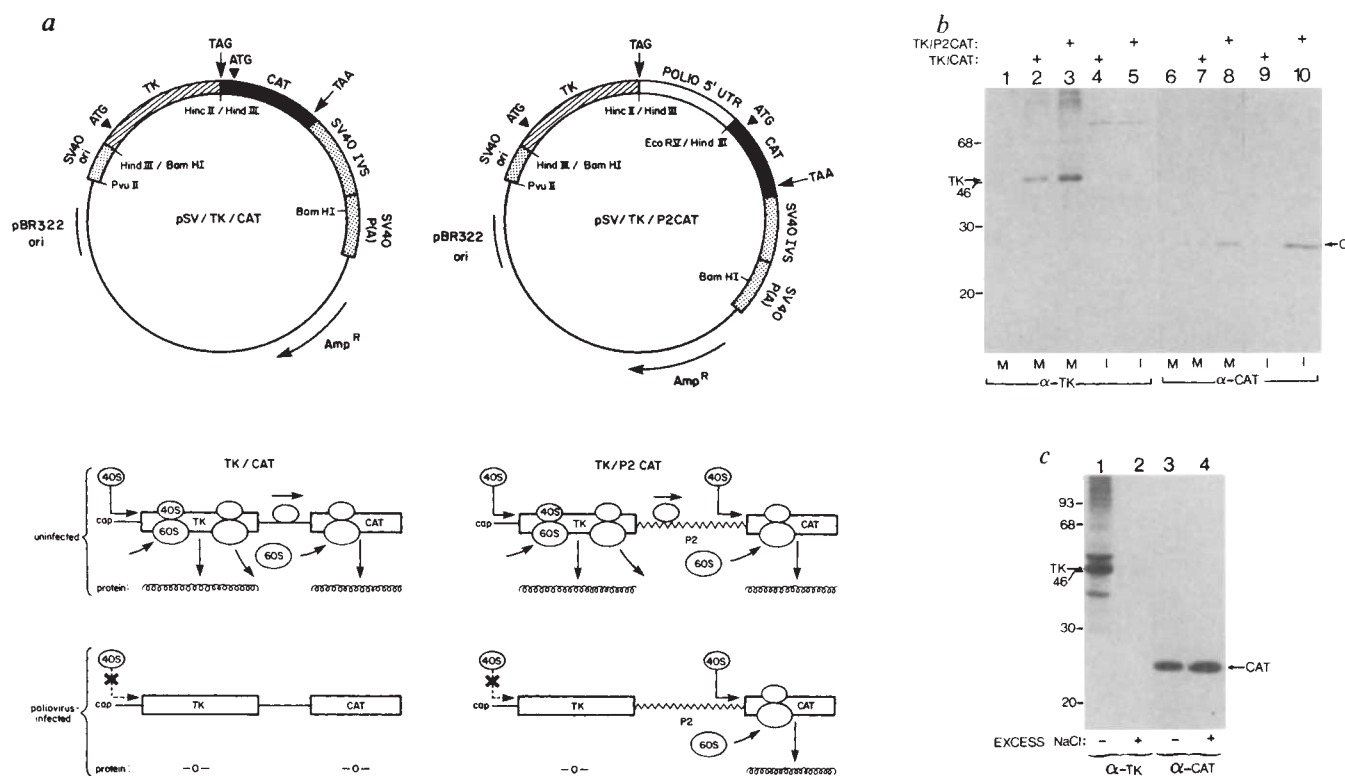


Fig. 1 *a*, Schematic diagram of bicistronic genes, showing the eukaryotic expression vector used for DNA transfections and experimental rationale. *b*, Immunoprecipitation of ³⁵S-labelled cell extracts prepared from mock (M) or poliovirus-infected (I) COS-1 cells. Cells were transfected with no DNA (lanes 1 and 6), pSV/TK/CAT (lanes 2, 4, 7, 9), or pSV/TK/P2CAT (lanes 3, 5, 8, 10). After 48 h, cells were infected with poliovirus type 1 (Mahoney strain, 50 plaque-forming units (PFUs) per cell) according to the method of Rose *et al.*⁴⁴. Four hours post-infection, cells were labelled with [³⁵S]methionine (100 µCi) in 1 ml α-minimal essential media (–Met) +10% dialysed fetal calf serum at 37 °C for 1 h. Lanes immunoprecipitated with anti-TK or anti-CAT antibodies are indicated at the bottom of the figure. *c*, Immunoprecipitation of ³⁵S-labelled cell extracts from mock or hypertonic treated COS-1 cells. Cells were transfected with pSV/TK/P2CAT. After 48 h the media was replaced with fresh media containing excess 190-mM NaCl and incubation was for 15 min followed by [³⁵S]methionine labelling as described for 1b.

Methods. Derivatives of pSV2 containing TK/CAT or TK/P2CAT were constructed from JP15/CAT and JP15/P2CAT (see Fig. 2a and ref. 22) which contain three *Bam*HI linkers within the TK 5' UTR. TK/CAT and TK/P2CAT were excised with *Bam*HI and the fragments were inserted into the *Hind*III/*Bam*HI sites of pSV2/CAT. COS-1 cells (seeded one day before transfection at 7.0×10^5 cells per 100-mm petri) were transfected with DNA (10 µg) using the DEAE-dextran method⁴⁵. Cell extracts were prepared⁴⁵, divided into two portions and immunoprecipitated with anti-CAT or anti-TK antibodies. Immunoprecipitates were analysed on 12% SDS/polyacrylamide gels that were treated with EN³HANCE, dried, and autoradiographed. Following immunoprecipitations, all cell extracts were put through a second round of immunoprecipitation to ensure there was no TK or CAT protein remaining in the supernatant.

directed by pSV/TK/P2CAT (as determined by immunoprecipitation) under physiological and hypertonic (excess 190-mM NaCl) conditions. Translation of TK is inhibited (~50-fold) when cells are grown under hypertonic conditions (compare lane 2 to 1). In contrast, CAT translation is even slightly stimulated under high salt conditions (compare lane 4 to 3). These results agree well with a model predicting independent and separate access points for ribosomes translating the TK and CAT cistrons.

An argument against internal ribosome binding could be that in poliovirus-infected cells, ribosomes do bind to the 5' end of the bicistronic mRNA, scan through all upstream AUGs, then for some unknown reason become 'activated' for AUG recognition by a signal in the poliovirus 5' UTR to initiate translation at the downstream CAT ORF. This possibility is extremely unlikely because of the overwhelming data showing that the CBP (cap binding protein) complex (eIF-4F) cannot bind to the cap structure of mRNAs following poliovirus-infection^{18,19}, thus excluding ribosome binding to the mRNA 5' end. Also, the results shown below (Fig. 2) argue strongly against this interpretation.

Translation *in vitro*

Because the *in vivo* experiments established internal initiation on poliovirus mRNA, we wished to reproduce these results in *in vitro* translation systems. If internal initiation could be observed *in vitro*, it would justify the use of translation extracts for the identification of putative *trans*-acting factors involved in internal initiation of translation and ultimately, elucidation of the molecular mechanism of this process. Translation studies were performed in a reticulocyte lysate using mRNAs synthesized *in vitro* in the SP6 transcription system (Fig. 2a). Translation of poliovirus mRNA is inefficient in this system²⁰ and we defined an inhibitory sequence for translation in reticulocyte lysates within the proximal 5' UTR (nucleotides 70–381; ref. 21). When the 5' UTR is juxtaposed 5' to the CAT coding sequence, translation of CAT is reduced ~200-fold (Fig. 2b, compare lane 3 to 2). Translation of TK mRNA gave a major polypeptide of 46K (Fig. 2c, lane 3), which corresponds to HSV-1 TK. The origin of the minor faster migrating polypeptides is not clear, but they are detectable only because the autoradiograph was greatly over exposed to detect the CAT protein.

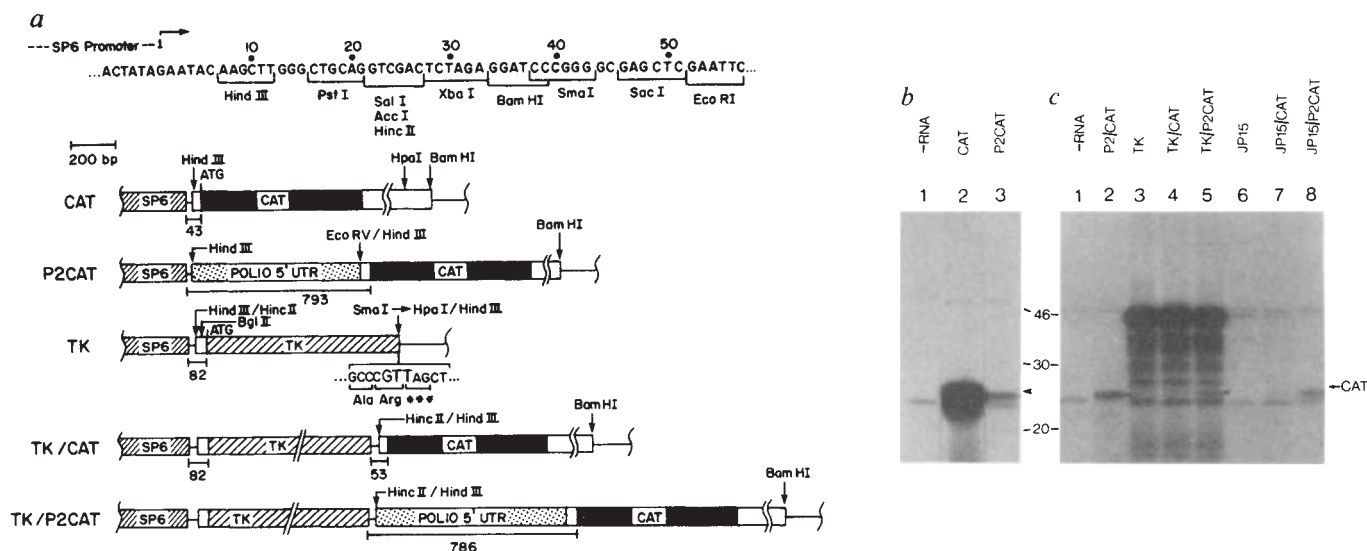


Fig. 2 *a*, Schematic representation of constructs for *in vitro* translation studies. The polylinker region of pSP64 is shown at the top of the diagram as a reference to the restriction sites used to create the various bicistronic genes described below. The right-angled arrow denotes the start and direction of transcription of SP6 polymerase. The numbers below the constructs represent distances in base pairs. The result of modifying the *TK* gene (pX1; see ref. 22) is shown below *TK*. The new artificially created stop codon in *TK* allowed us to easily manipulate intercistronic distance. The large letters represent bases derived from the *Hpa*I linker. The blank boxes denote 5' and 3' UTRs, and the thin line represents sequences from pSP64. *b*, SDS-polyacrylamide gel electrophoresis of proteins synthesized in rabbit reticulocyte lysate programmed with *in vitro* transcribed mRNA. Transcription of *Sma*I linearized plasmids using SP6 polymerase was performed as previously described²². Translations were done in 12- μ l reaction mixtures containing 0.2 μ g mRNA at 30 °C for 1 h⁴⁶. Four microlitres of mixture was withdrawn after incubation, analysed on 12% SDS-polyacrylamide gels that were treated with EN³HANCE, dried and autoradiographed. **Methods.** CAT and P2CAT have been previously described¹⁴. *TK* was derived from pX1 (ref. 22) by replacing the *Sma*I site, which resides within the seventh codon before the termination codon, with a *Hpa*I linker (5'GTTAAC3'). This derivative was subcloned into pSP64 by restricting with *Hinc*II (which cleaves within the *TK* 5' UTR) and *Hpa*I and inserting into the blunted *Hind*III site of pSP64. This generates an in-frame amber termination codon. TK/CAT and TK/P2CAT were created by first digesting *TK* with *Hinc*II and *Bam*HI. The CAT and P2CAT fragments isolated from CAT and P2CAT by restricting with *Hind*III were blunted with Klenow, followed by *Bam*HI digestion and ligated to *TK*. JP15 derivatives containing three *Bam*HI linkers in the non-coding region of *TK* (in *c*) were similarly constructed by using pXJP15 described in ref. 22.

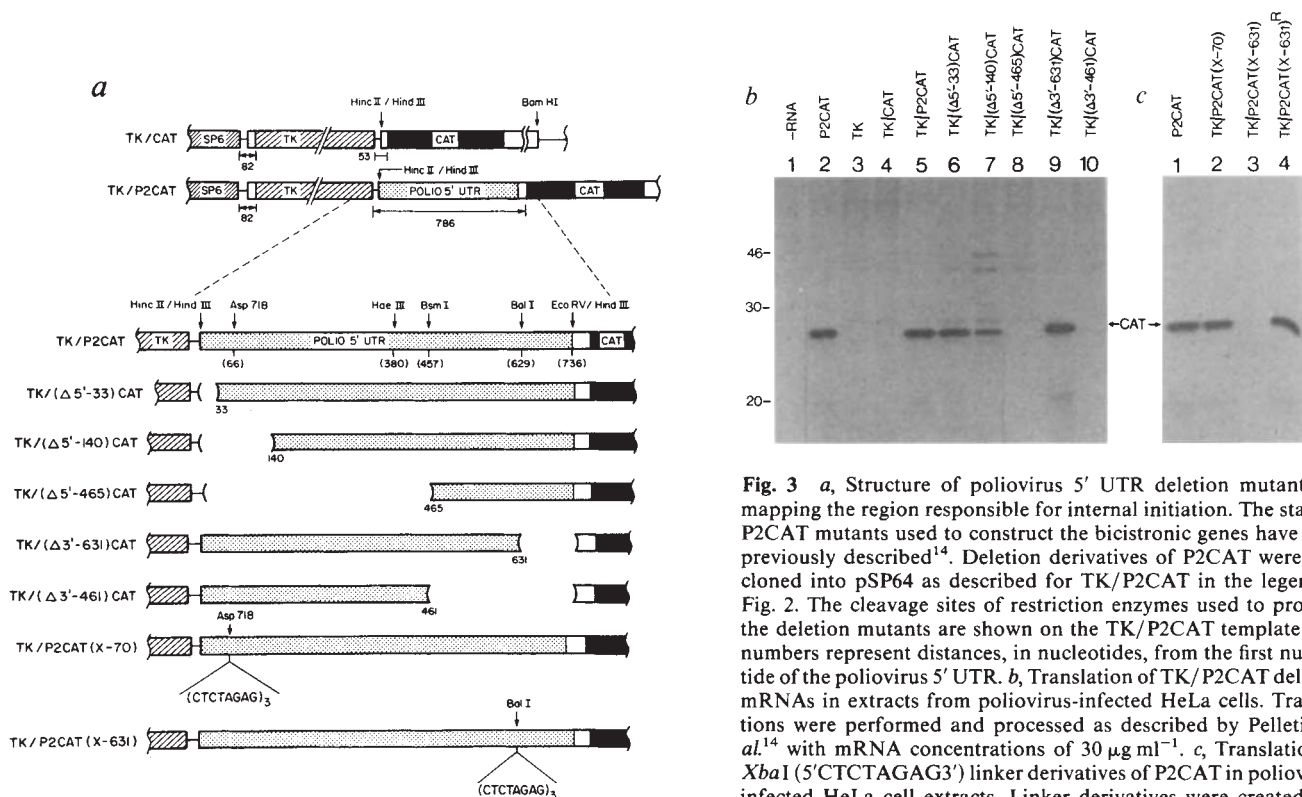


Fig. 3 *a*, Structure of poliovirus 5' UTR deletion mutants for mapping the region responsible for internal initiation. The starting P2CAT mutants used to construct the bicistronic genes have been previously described¹⁴. Deletion derivatives of P2CAT were subcloned into pSP64 as described for TK/P2CAT in the legend to Fig. 2. The cleavage sites of restriction enzymes used to produce the deletion mutants are shown on the TK/P2CAT template. The numbers represent distances, in nucleotides, from the first nucleotide of the poliovirus 5' UTR. *b*, Translation of TK/P2CAT deletion mRNAs in extracts from poliovirus-infected HeLa cells. Translations were performed and processed as described by Pelletier *et al.*¹⁴ with mRNA concentrations of 30 μ g ml⁻¹. *c*, Translation of *Xba*I (5'CTCTAGAG3') linker derivatives of P2CAT in poliovirus-infected HeLa cell extracts. Linker derivatives were created and characterized as described²². Transcriptions and translations were performed and processed as described in ref. 14.

Translation of TK/CAT gave an identical pattern of proteins to that of TK mRNA, indicating that the downstream CAT cistron was not used in this construct (compare lane 4 to 3). But when TK/P2CAT was translated, CAT was synthesized (indicated by arrowhead; lane 5), most probably by internal initiation, as was shown *in vivo* (Fig. 1). This contention is further supported by translation of insertion derivatives of TK and TK/CAT (termed JP15 and JP15/CAT) that contain three tandem repeats of *Bam*HI linkers (5'CCGGATCCGG3') having the potential to form a stable secondary structure within the TK 5' UTR²². The translation of mRNAs from these constructs is strongly inhibited in a reticulocyte lysate²². Translation of JP15 and JP15/CAT mRNAs was very inefficient as anticipated (lanes 6 and 7)²². Remarkably, translation of the downstream cistron of CAT in JP15/P2CAT was not affected by the secondary structure upstream of the TK cistron (compare lane 8 to 5). These *in vitro* results mimic those obtained *in vivo* (Fig. 1) and provide evidence for internal initiation *in vitro*.

Mapping of ribosome binding site

A series of deletion mutants within the poliovirus 5' UTR intercistronic region were constructed to map the region required for internal initiation of translation (Fig. 3a). Translations were performed in extracts from poliovirus-infected HeLa cells (Fig. 3b), which do not permit the translation of cap-dependent mRNAs^{18,23}. Translation of P2CAT mRNA was efficient in this extract (lane 2) and translation of the downstream CAT cistron in TK/P2CAT was as efficient (compare lane 5 to 2). As was shown *in vivo* under conditions of poliovirus infection, CAT was not expressed from the bicistronic construct TK/CAT (lane 4). As expected, there was no expression from the upstream TK cistron in these extracts for any of the constructs. Deletion of 33 nucleotides from the 5' UTR had no effect (lane 6), whereas deletion of 140 nucleotides reduced CAT translation threefold (lane 7). A further deletion of 465 nucleotides from the 5' poliovirus UTR completely abolished internal initiation (lane 8). A deletion removing 100 nucleotides from the 3' proximal UTR [TK/(Δ 3'-631)/CAT] slightly stimulated (twofold; lane 9) translation, showing that the 3' proximal nucleotides of the poliovirus 5' UTR are not required for internal initiation. But a further deletion from the 3' end extending to nucleotide 461 completely abolished CAT translation (lane 10).

These results indicate that sequences between nucleotides 140 and 630 of the 5' UTR contain the internal ribosome-binding site. Similar results were obtained when the constructs were translated in uninfected extracts, except that the appearance of TK products was observed. Similar results were also observed when JP15 was used as the first cistron for the constructs shown in Fig. 3a and translations were performed in poliovirus-infected or uninfected HeLa extracts, except that TK production was reduced in uninfected HeLa extracts (data not shown). This finding is consistent with our previous result that an internal sequence of the poliovirus UTR confers cap-independent translation when fused 5' to a heterologous mRNA¹⁴.

The results suggest that ribosomes bind first to an internal sequence of the poliovirus 5' UTR and are then translocated, presumably by scanning, to the initiator AUG of the CAT ORF. This interpretation is supported by the experiment shown in Fig. 3c. Three *Xba*I linkers (5'CTCTAGAG3'), having the potential to form a hairpin structure in the mRNA ($\Delta G^\circ = -30$ kcal mol⁻¹) and thus inhibit translation initiation²², were introduced into the 5' proximal region (at position 70) or at the 3' distal region (at position 631) of the poliovirus 5' UTR. Insertion of the *Xba*I linkers at position 70 had no effect on translation of CAT (compare lane 2 to 1), whereas insertion at position 631 dramatically inhibited P2CAT translation (compare lane 3 to 1). The observed inhibition is not due to a second site mutation because removal of the *Xba*I linkers [TK/P2CAT(X-631)^R] restored translation to levels observed with P2CAT mRNA (compare lane 4 to 1). Thus, ribosomes must bind to

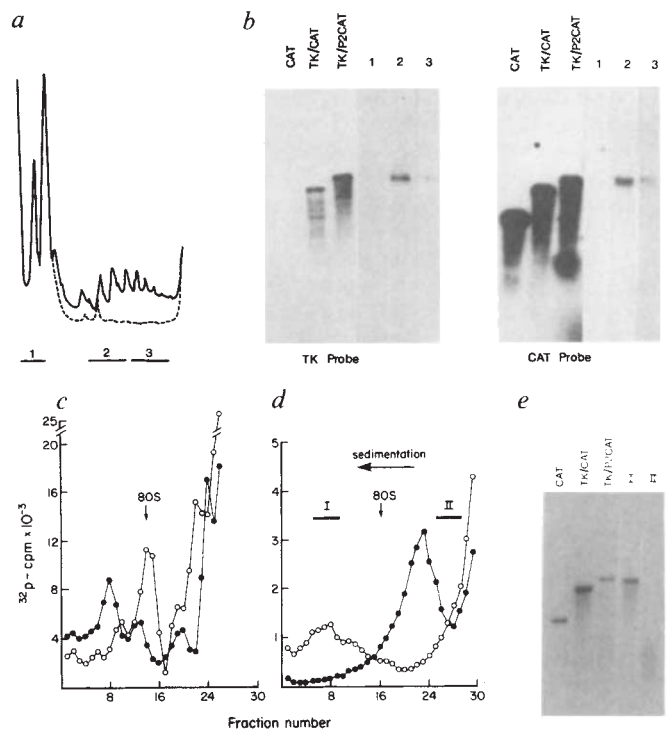


Fig. 4 Integrity of TK/P2CAT *in vivo* and *in vitro*. **a**, Polysome profiles from uninfected COS-1 cells (black line) and poliovirus-infected, TK/P2CAT-transfected COS-1 cells (dashed lines). **b**, Northern blot analysis of mRNA fractions purified from portions of the polysome gradient denoted in **a**. The probes used consisted of the ³²P-labelled *Hinc*II-*Sma*I TK fragment from pX1 (ref. 22) and the *Hind*III-*Bam*HI CAT fragment from pSV2/CAT⁴⁵. The lanes indicated by CAT, TK/CAT and TK/P2CAT contained SP6 generated transcripts used for size markers. Polysomes were prepared fractionated and analysed as described by Katze *et al.*⁴⁷. **c-e**, Ribosome binding and integrity of mRNAs containing the poliovirus 5' UTR. **c** and **d**, α -³²P-ATP-labelled CAT, P2CAT and TK/P2CAT (2×10^5 c.p.m.; 1 μ g for the experiment shown in **c**, and 6×10^5 c.p.m.; 3 μ g for the experiment in **d**) were incubated in 32 μ l of uninfected (**c**) or poliovirus-infected (**d**) HeLa extract at 30 °C for 10 min as described by Weber *et al.*⁴⁸. Initiation complexes were formed and analysed on glycerol gradients⁴⁸. Fractions were collected and radioactivity was determined by scintillation counting (**c**) or by Cherenkov counting (**d**). The horizontal bars designated I and II indicate the fractions pooled for mRNA analysis. **e**, The pooled fractions from experiment **b** were treated with proteinase K (400 μ g ml⁻¹) at 37 °C for 30 min, phenol extracted and ethanol precipitated. The transcripts were analysed on a 1.2% formaldehyde agarose gel.

the poliovirus 5' UTR upstream of position 631, but downstream of position 70, and subsequently reach the AUG initiator at position 745 to initiate translation.

Translated mRNA is intact

Our results can best be explained by ribosomes binding internally to the poliovirus 5' UTR. We wished, however, to exclude the unlikely possibility that nucleolytic cleavage of the bicistronic TK/P2CAT mRNA generated fragmented mRNAs capable of translation in poliovirus-infected cells. To do this, we analysed the integrity of TK/P2CAT mRNA obtained from polysomes isolated from poliovirus-infected cells. The polysome profile from uninfected and poliovirus-infected COS-1 cells is shown in Fig. 4a. Poliovirus infection caused a sharp decrease in the size and amount of polysomes²³. A northern blot of mRNA isolated from different fractions of the polysome gradient (Fig. 4a), probed with a ³²P-TK or ³²P-CAT DNA fragment, shows that the mRNA is intact in poliovirus-infected, TK/P2CAT

transfected COS-1 cells (Fig. 4b). Most of the mRNA is present on small polysomes, consistent with only the CAT portion of the mRNA being translated. The translation of the CAT cistron in TK/P2CAT mRNA could not have been achieved by translation of a fragmented mRNA.

We also examined the integrity of TK/P2CAT mRNA in initiation complexes formed in extracts from poliovirus-infected cells. Figure 4c shows a glycerol gradient sedimentation analysis of initiation complexes formed with CAT and P2CAT mRNA in extracts from uninfected HeLa cells. CAT mRNA formed mainly an 80S initiation complex as expected. But P2CAT mRNA was engaged mainly in disomes, indicating that most P2CAT molecules bound two ribosomes, one presumably at the CAT AUG initiator and a second ribosome upstream in the poliovirus mRNA 5' noncoding region. TK/P2CAT mRNA formed a similar size complex in extracts from poliovirus-infected HeLa cells (Fig. 4d). Initiation complexes were not formed in the presence of EDTA (Fig. 4d) or at 4 °C (data not shown), consistent with them being translation specific complexes. Also, TK/CAT mRNA which does not contain any poliovirus RNA sequence, did not form initiation complexes in poliovirus-infected HeLa cell extracts (data not shown). Messenger RNA from initiation complexes (region I in Fig. 4d) and from the top of the gradient (region II in Fig. 4d) was extracted and analysed on a formaldehyde agarose gel (Fig. 4e). The RNA present in the initiation complex is intact, whereas the RNA from the top of the gradient is mostly degraded. We conclude that TK/P2CAT mRNA entering initiation complexes is intact, thus excluding the possibility that fragmentation of the mRNA is a prerequisite for translation of this mRNA.

Discussion

In this paper we present evidence that eukaryotic ribosomes can bind internally to the 5' noncoding region of poliovirus RNA. This finding is consistent with several peculiarities of poliovirus translation including the absence of a cap structure^{1,2,4,5}, expression under conditions where the cap binding protein complex (eIF-4F) is inactivated^{15,19,23} and the presence of an internal sequence in the 5' noncoding region which confers cap-independent translation to the mRNA¹⁴. It is not the mere difference in length of the intercistronic region which allows TK/P2CAT to be expressed in infected extracts because insertion of 445 nucleotides from the mouse *c-myc* 5' UTR as the intercistronic region did not allow translation of CAT in mock-infected HeLa extracts (unpublished observations). Also, TK/(Δ3'-461)CAT, which had an intercistronic spacer of 511 nucleotides did not translate in poliovirus-infected extracts (Fig. 3b, lane 10). These results support the contention that eukaryotic ribosomes are capable of binding internally on certain mRNAs. The strongest evidence against this possibility is the finding that eukaryotic ribosomes are unable to bind to circular RNA, whereas prokaryotic ribosomes can^{13,24}. Unfortunately, these studies suffer from several disadvantages. The RNA fragments were small (<80 nucleotides) and most probably lacked important signals for internal ribosome binding¹³ or their circularization could have sterically prevented ribosome binding^{13,24}.

It is possible that poliovirus RNA constitutes an extreme exception for internal ribosome binding and that the majority of other eukaryotic mRNAs do not have the necessary signal sequences. Several *in vitro* results with other mRNAs have been interpreted as indication that translation initiation at internal sequences can occur without scanning from the 5' end. The most relevant study is of another member of the picornavirus family, encephalomyocarditis (EMC) virus: complementary DNA fragments to the 5' proximal 450 nucleotides of EMC virus RNA had no effect on the synthesis of the EMC polyprotein²⁵. Moreover, a recent study showed directly that an internal sequence in the 5' noncoding region of EMC virus RNA directs internal binding of ribosomes *in vitro*²⁶. Because all picornavirus RNAs have similar translational characteristics in that they are

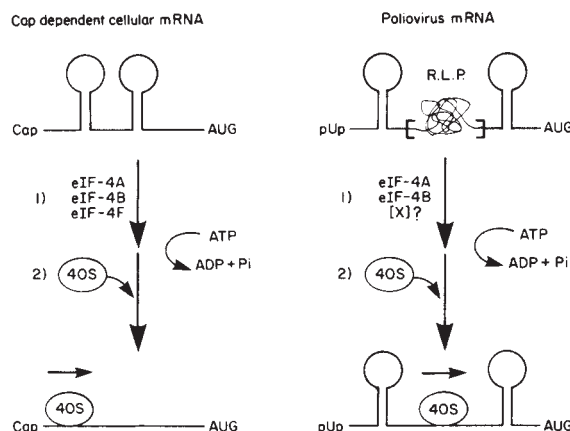


Fig. 5 Models for cap-dependent and internal binding of ribosomes to mRNA. The mechanism for cap-dependent initiation is based on many lines of evidence as reviewed by Edery *et al.*⁴⁹. The internal binding mechanism might be facilitated by additional factors (denoted by X) and mRNA signals (depicted by RLP (Ribosome Landing Pad)). The requirement for the hydrolysis of ATP in internal binding is as yet ill-defined.

cap independent, it is most probable that all of them are initiated by direct internal binding of ribosomes. Results from *in vitro* hybrid-arrest translation experiments were also interpreted as evidence for internal initiation on other viral mRNAs including the NS (P) mRNA of vesicular stomatitis virus (VSV)²⁷; adenovirus-2 polymerase mRNA²⁸ and infectious pancreatic necrosis virus segment A²⁹.

Internal initiation on cellular mRNAs? There are several potential and interesting candidates for internal initiation among cellular mRNAs—for example, the yeast GCN4 mRNA which contains four mini-cistrons (2–3 amino acids) upstream of the GCN4 long ORF^{30,31}. These mini-cistrons regulate translation from the downstream GCN4 ORF. The 5' proximal upstream ORF acts as a positive regulator under starvation conditions^{32,33}. It is possible that under these conditions translation of the first mini-cistron facilitates internal initiation at the GCN4 AUG. Several other candidates for internal initiation are mRNAs that possess unusually long 5' noncoding regions containing multiple AUGs. For example, one of the transcripts of the *Antennapedia* gene of *Drosophila melanogaster* contains a 5' noncoding region of ~1,100 nucleotides with 15 AUG codons³⁴. One of the human *c-abl* transcripts (type 1b) has a 5' noncoding region of 1,200 nucleotides containing 13 AUG codons³⁵ and the hamster *HMG CoA reductase* gene produces a family of mRNAs with variable 5' UTR length, the longest containing 670 nucleotides and eight AUG codons upstream of the initiation site for translation³⁶. Heat-shock mRNAs constitute another set of mRNAs that might initiate translation by internal ribosome binding. Internal sequences of *Drosophila* heat-shock mRNAs 5' UTR are required for translation under heat-shock conditions^{37,38}. Moreover, there is some evidence that the CBP complex (eIF-4F) may be rendered inactive at elevated temperatures³⁹, suggesting that translation of heat-shock mRNAs is less dependent on the cap structure than the bulk of cellular mRNAs. This is consistent with the finding that heat-shock mRNAs are more resistant than the bulk of cellular mRNAs to the shut-off of host protein synthesis after poliovirus infection⁴⁰.

Model for ribosome internal binding. What is the molecular mechanism that explains internal binding of ribosomes to eukaryotic mRNAs? There is considerable evidence that the cap structure of eukaryotic mRNAs mediates the melting of regions of secondary structure in the 5' UTR through the action of the cap-binding protein complex (eIF-4F)^{41,42}, thus facilitating 40S ribosomal subunit binding. This melting activity is dependent

on two other initiation factors eIF-4A and eIF-4B^{42,43}. Poliovirus RNA translation proceeds by a cap-independent mechanism and most probably does not require eIF-4F, but does require eIF-4A and eIF-4B. The model we propose for poliovirus translation is schematically shown in Fig. 5. According to our proposed model, eIF-4A in conjunction with eIF-4B, binds to an internal sequence of the mRNA. The nature of the sequence, recognized by eIF-4A and eIF-4B, which we have termed the ribosome landing pad (RLP) is not known. But an intriguing possibility is that this sequence is naturally devoid of secondary structure or is melted by an unknown factor(s) (X in Fig. 5). ATP-dependent binding of eIF-4A to unstructured RNA has previously been documented⁴³. This would then signal the binding of ribosomes. The importance of single-stranded regions does not exclude, however, the possibility that secondary and

tertiary structure of the RLP region serve as recognition signals for factors involved in ribosome binding. The size of the region required for internal binding (several hundred nucleotides; Fig. 3b) suggests that secondary and tertiary structure or additional factors (denoted as X in Fig. 5) are required for efficient internal ribosome binding. The system we have described should allow the testing of putative 5' UTRs of cellular mRNAs which may be involved in internal ribosome binding.

We thank Drs J. Smiley and W. Summers for their gift of anti-HSV-1 TK rabbit polyclonal antibody, Dr C. Gorman for her gift of anti-CAT antibody, Drs V. Racaniello and G. Kaplan for poliovirus complementary DNA clones, and I. Edery and N. Parkin for comments on the manuscript. This work was supported by the Medical Research Council of Canada.

Received 22 February; accepted 10 May 1988.

- Lee, Y. F., Nomoto, A., Detjen, B. M. & Wimmer, E. *Proc. natn. Acad. Sci. U.S.A.* **74**, 59–63 (1977).
- Flanagan, J. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 961–965 (1977).
- Ambros, V., Petterson, R. F. & Baltimore, D. *Cell* **15**, 1439–1446 (1978).
- Hewlett, M. J., Rose, J. K. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 327–330 (1976).
- Nomoto, A., Lee, Y. F. & Wimmer, E. *Proc. natn. Acad. Sci. U.S.A.* **73**, 375–380 (1976).
- Kitamura, N. *et al. Nature* **291**, 547–553 (1981).
- Racaniello, V. R. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4887–4891 (1981).
- Toyoda, H. *et al. J. molec. Biol.* **174**, 561–585 (1984).
- Kozak, M. *Microbiol. Rev.* **47**, 1–45 (1983).
- Liu, C.-C., Simonson, C. C. & Levinson, A. D. *Nature* **309**, 82–85 (1984).
- Hughes, S., Mellstrom, K., Kosik, E., Tamanoi, F. & Brugge, J. *Molec. cell. Biol.* **4**, 1738–1746 (1984).
- Kozak, M. *Cell* **47**, 481–483 (1986).
- Kozak, M. *Nature* **280**, 82–85 (1979).
- Pelletier, J., Kaplan, G., Racaniello, V. R. & Sonenberg, N. *Molec. cell. Biol.* **8**, 1103–1112 (1988).
- Etchison, D. *et al. J. biol. Chem.* **257**, 14806–14810 (1982).
- Nuss, D. L., Opperman, H. & Koch, G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1258–1262 (1975).
- Carrasco, L. & Smith, A. E. *Nature* **264**, 807–809 (1976).
- Lee, K. A. W. & Sonenberg, N. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3447–3451 (1982).
- Lee, K. A. W., Edery, I. & Sonenberg, N. *J. Virol.* **54**, 515–524 (1985).
- Shih, D. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 5807–5811 (1978).
- Pelletier, J., Kaplan, G., Racaniello, V. R. & Sonenberg, N. *J. Virol.* **62**, 2219–2227 (1988).
- Pelletier, J. & Sonenberg, N. *Cell* **40**, 515–526 (1985).
- Sonenberg, N. *Adv. Virus Res.* **33**, 175–204 (1987).
- Konarska, M., Filipowicz, W., Domdey, H. & Gross, H. J. *Eur. J. Biochem.* **114**, 221–227 (1981).
- Shih, D. S., Park, I.-W., Evans, C. L., Jaynes, J. M. & Palmenberg, A. C. *J. Virol.* **61**, 2033–2037 (1987).
- Jang, S. K. *et al. J. Virol.* (in the press).
- Herrman, R. C. *J. Virol.* **58**, 797–804 (1986).
- Hassin, D., Korn, R. & Horwitz, M. S. *Virology* **155**, 214–224 (1986).
- Nagy, E., Duncan, R., Krell, P. & Dobos, P. *Virology* **158**, 211–217 (1987).
- Hinnebusch, A. G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6442–6446 (1984).
- Thireos, G., Penn, M. D. & Greer, H. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5096–5100 (1984).
- Mueller, P. P. & Hinnebusch, A. G. *Cell* **45**, 201–207 (1986).
- Tzamaras, P., Alexandraki, D. & Thireos, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4849–4853 (1986).
- Stroehrer, V. L., Jorgensen, E. M. & Garber, R. L. *Molec. cell. Biol.* **6**, 4667–4675 (1986).
- Bernards, A., Rubin, C. M., Westbrook, C. A., Paskind, M. & Baltimore, D. *Molec. cell. Biol.* **7**, 3231–3236 (1987).
- Reynolds, G. A., Goldstein, J. L. & Brown, M. S. *J. biol. Chem.* **260**, 10369–10377 (1985).
- McGarry, T. J. & Lindquist, S. *Cell* **72**, 903–911 (1985).
- Hultmark, D., Klemenz, R. & Gehring, W. J. *Cell* **44**, 429–438 (1986).
- Panniers, R., Stewart, E. B., Merrick, W. C. & Henshaw, E. C. *J. biol. Chem.* **260**, 9648–9653 (1985).
- Munoz, A., Alonso, M. A. & Carrasco, L. *Virology* **137**, 150–159 (1984).
- Edery, I., Lee, K. A. W. & Sonenberg, N. *Biochemistry* **23**, 2456–2462 (1984).
- Ray, B. *et al. J. biol. Chem.* **260**, 7651–7658 (1985).
- Abramson, R. D. *et al. J. biol. Chem.* **262**, 3826–3832 (1987).
- Rose, J. K., Trachsel, H., Leong, K. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2732–2736 (1978).
- Gorman, C. in *DNA Cloning* Vol. 2 (ed. Glover, D. M.) (IRL Press, Oxford, 1985).
- Pelham, H. P. B. & Jackson, R. J. *Eur. J. Biochem.* **67**, 247–256 (1976).
- Katze, M. G., DeCorato, D. & Krug, R. M. *J. Virol.* **60**, 1027–1039 (1986).
- Weber, L. A., Simili, M. & Baglioni, C. *Meth. Enzym.* **60**, 351–360 (1979).
- Edery, I., Pelletier, J. & Sonenberg, N. in *Translational Regulation of Gene Expression* (ed. Ilan, J.) 335–366 (Plenum, New York, 1985).

LETTERS TO NATURE

Discovery of a quadruply lensed quasar: the 'clover leaf' H1413+117

P. Magain*, J. Surdej†, J.-P. Swings†, U. Borgeest‡, R. Kayser‡, H. Kühr§, S. Refsdal‡ & M. Remy||

* European Southern Observatory, Karl-Schwarzschild Strasse 2, D-8046 Garching bei München, FRG

† Institut d'Astrophysique, Université de Liège, Avenue de Cointe 5, B-4200 Cointe-Ougrée, Belgium

‡ Hamburger Sternwarte, Gojenbergsweg 112, D-2050 Hamburg 80, FRG

§ Max-Planck-Institut für Astronomie, Königstuhl, D-6900 Heidelberg 1, FRG

|| European Southern Observatory, Casilla 19001, Santiago 19, Chile

In November 1986 we began an optical search for examples of gravitational lensing in a sample of highly luminous quasars (HLQs, $M_v < -29$), with the aims of improving our knowledge of the quasar luminosity function, studying the dark matter content of the Universe, and redetermining some important cosmological parameters. This survey has found one new case of lensing^{1,2} and the general implications of the search have been summarized³. Here we report the discovery of a second gravitational lens system

in the broad absorption line quasar H1413+117 (refs 4–6). Four images of comparable brightness are seen, separated by ~ 1 arcsec. Spectra obtained of two of the images are identical apart from the presence of sharp absorption lines in one component, which are presumably due to gas clouds along the line of sight. The unique configuration of the images, together with the fairly rare occurrence of this type of quasar, makes it incontrovertible that this is a lensed system, not a cluster of quasars, and this second discovery made by imaging bright quasars establishes the power of the method for finding systems with small separations.

H1413+117, with a redshift of 2.55 and an apparent visual magnitude of 17, is one of the brightest members of the class of broad absorption line (BAL) quasars. The latter represent a few percent of the total number of quasars and display broad and deep absorptions on the short wavelength side of some of their emission lines^{7,8}. The spectrum of H1413+117, which has already been studied in detail^{4–6}, shows particularly smooth and deep lines of the P Cygni type. In addition, two narrow absorption line systems have been identified, at redshifts 1.66 and 2.07. These are attributed to intervening gas clouds, possibly associated with the lens(es).

Our images of H1413+117 were obtained with a charge-coupled device (CCD) camera at the Cassegrain focus of the ESO (European Southern Observatory)/MPI (Max Planck Institut) 2.2 m telescope at La Silla. A double density RCA CCD

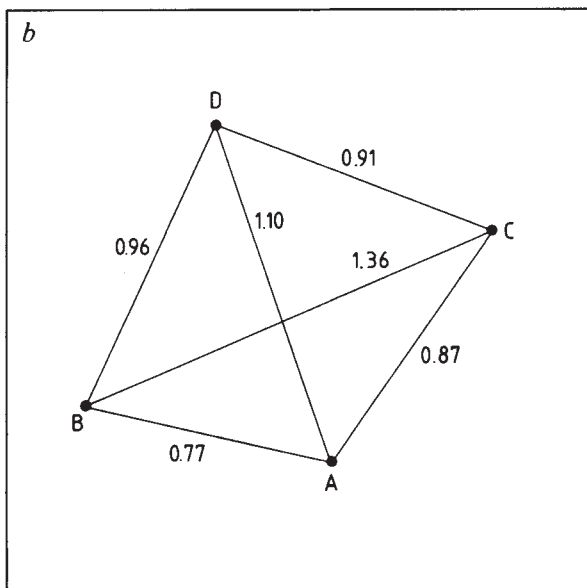
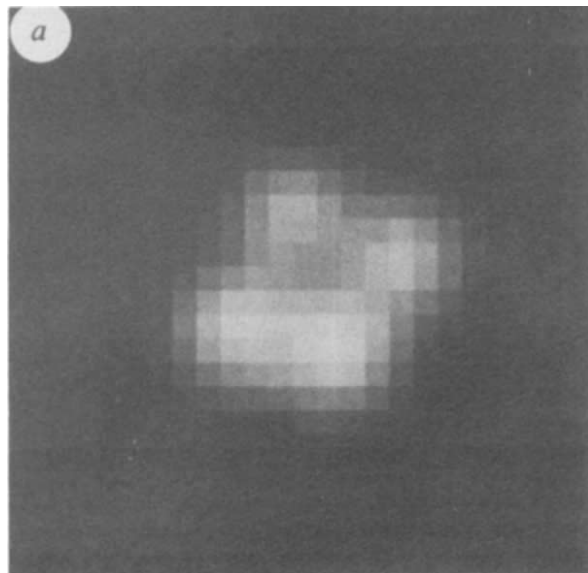


Fig. 1 *a*, The central part of a CCD frame of H1413 + 117 obtained with the ESO/MPI 2.2 m telescope through a Bessel R filter on 8 March 1988. *b*, Geometry of the system. Distances between components are indicated in seconds of arc. North is at the top and East to the left. Courtesy of ESO.

was used, each $15\ \mu\text{m}$ pixel corresponding to $0.176\ \text{arcsec}$ on the sky. Good seeing conditions ($0.8\ \text{arcsec}$ FWHM) allowed us to resolve the quasar image into four components (Fig. 1*a*). The positions and magnitudes of these four images were determined by simultaneously fitting double gaussian profiles, whose parameters had been determined from neighbouring stellar images. These are listed in Table 1, and Fig. 1*b* sketches the geometry of the system. Four CCD frames have been obtained and analysed in this way and give the same results within 2%, which is thus an estimate of the accuracy of our measurements. The mean separation between the images amounts to one arcsec, making this system the most compact case of gravitational lensing known so far.

Very good evidence that H1413 + 117 is gravitationally lensed comes from the peculiarity of its spectrum⁴⁻⁶, with broad absorptions reaching nearly zero residual intensity. If any of the four images were to correspond to another object, it is highly improbable that this object would have deep and broad absorption

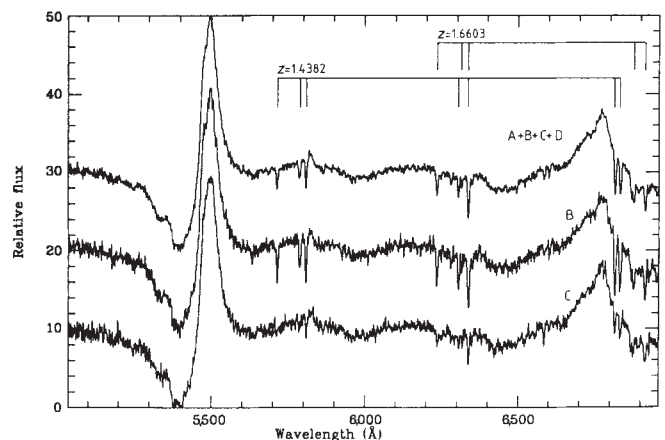


Fig. 2 Low dispersion spectra of components B and C, and of the whole system. The ordinate scale refers to the lower spectrum; the others are displaced vertically by 10 and 20 units. The two identified narrow absorption line systems are indicated.

lines at exactly the same wavelengths. Thus, the light from such an object would fill in the absorptions, in contradiction with the observations.

Further evidence for gravitational lensing is provided by our own spectra. These were obtained with the ESO faint object spectrograph and camera (EFOSC) (ref. 9) at the Cassegrain focus of the ESO 3.6-m telescope at La Silla. The instrument was rotated to roughly align the slit along the line joining the B and C images (position angle 120°). The O150 grism⁹ was used, giving a spectral resolution of $\approx 7\ \text{\AA}$ in the range 5,000 to 7,000 $\text{\AA}$. Three 30 min spectra were obtained and co-added. The mean seeing of $0.9\ \text{arcsec}$ valued us to separate the spectra of the B and C components, although some contamination from A and D could not be avoided. These spectra are reproduced in Fig. 2 with the integrated spectrum (all four components). Within the limits of the accuracy of our data, the spectra of B and C are identical as far as the quasar features are concerned (continuum, emission and broad absorption lines), supporting the interpretation in terms of a gravitational lens. They are also identical to the whole integrated spectrum, indicating that components A and D probably also have the same spectrum. The emission line redshifts of B and C are identical at $z = 2.551$.

The only difference between the spectra of B and C lies in the narrow absorption lines, which are much stronger in the spectrum of B. The weak lines in the C spectrum may result from some contamination by the light of the other images. These absorption lines belong to two systems, at redshifts 1.4382 ± 0.0002 (new) and 1.6603 ± 0.0003 . The rather large redshift difference between these systems implies that the corresponding clouds are probably not associated.

Although the high degree of symmetry of this system would be easier to explain if the lensing object is located in between the four images, the presence of narrow absorption lines in the spectrum of the B image rather suggests that it lies in front of this component. But in the absence of any more detailed information on the lens, we have applied a simple model consisting of an axially symmetric galaxy lying on the optical axis, to which

Table 1 Geometry of the system

Component	$\Delta\alpha$	$\Delta\delta$	Δm_R
A	0.00	0.00	0.00
B	0.75	0.17	0.15
C	-0.50	0.71	0.30
D	0.35	1.04	0.40

$-\Delta\alpha, \Delta\delta$ = coordinates measured with respect to image A, in seconds of arc. Δm_R = magnitude difference with respect to image A, in filter R.

Table 2 Estimates of deflector mass and upper limit on time delay

Lens redshift	$M/10^{11} M_{\odot}$	$\Delta t_{\max}$ (days)
0.50	1.0	15
1.00	2.5	50
1.44	4.8	120
1.66	6.8	170
2.07	15.3	350

a very small asymmetric disturbance is added^{10,11}. (Note that with such a transparent and non-singular lens, there should be a central fifth image, but it would be much too faint to be seen). The estimated mass M of the lensing galaxy is given in Table 2 as a function of the adopted redshift, including the redshifts of the known narrow absorption line systems. All data in this Table are computed assuming a value of $75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ for the Hubble constant H_0 and a deceleration parameter $q_0 = 0$. The calculated masses are consistent with common estimates of galaxy masses. Rough upper limits $\Delta t_{\max}$ on the delay between the travel times of two opposite images are also listed in Table 2. More probable values are at least a factor of two smaller.

The expected time delays are short enough to be measured in one observing season. But, particularly for small lens redshifts, they may be even shorter than the variability timescale of the quasar itself, in which case they would be difficult to determine at all. On the other hand, such short time delays present the advantage that micro-lensing variability would be easier to distinguish from intrinsic variability. As there are four images close to the centre of the (hypothetical) lens galaxy, we expect the total probability for micro-lensing to be higher in this system than in any of the other known cases of gravitational lensing.

Received 25 April; accepted 10 June 1988.

1. Surdej, J. *et al.* *Nature* **329**, 695–696 (1987).
2. Surdej, J. *et al.* *Astr. Astrophys.* **198**, 49–60 (1988).
3. Surdej, J. *et al.* *Publ. Astron. Soc. Pac. Suppl.*, *ESO Preprint no. 571* (Proc. NOAO Workshop on Optical Surveys for Quasars, in the press).
4. Hazard, C., Morton, D. C., Terlevich, R. & McMahon, R. *Astr. Astrophys. J.* **282**, 33–52 (1984).
5. Dekker, H. & Bokkenberg, A. *Mon. Not. R. Astr. Soc.* **211**, 813–831 (1984).
6. Turnshek, D. A., Foltz, C. B., Grillmair, C. J. & Weymann, R. J. *Astr. Astrophys. J.* **325**, 651–670 (1988).
7. Weymann, R. J., Carswell, R. F. & Smith, G. A. *Rev. Astr. Astrophys.* **19**, 41–76 (1981).
8. Surdej, J. & Hutsemekers, D. *Astr. Astrophys.* **177**, 42–50 (1987).
9. Dekker, H. & D'Odorico, S. *European Southern Observatory Operating Manual no. 4* (1985).
10. Refsdal, S. *Mon. Not. R. Astr. Soc.* **132**, 101–111 (1966).
11. Borgeest, U. & Refsdal, S. *Astr. Astrophys.* **141**, 318–332 (1984).

Carbon monoxide in supernova 1987A

J. Spyromilio, W. P. S. Meikle,
R. C. M. Learner & D. A. Allen*

Blackett Laboratory, Imperial College of Science and Technology,
Prince Consort Road, London SW7 2BZ, UK

* Anglo-Australian Observatory, PO Box 296, Epping,
NSW 2121, Australia

The identification of carbon monoxide in the infrared spectra of SN1987A (refs 1–4), which exploded on 23 February 1987, represents the first-ever detection of a molecule in a supernova. In infrared spectra obtained with the Anglo-Australian Telescope, first overtone emission from CO was observed near $2.3 \mu\text{m}$ as early as day 112 after the explosion, and was clearly detectable in all subsequent spectra. Spectra of the CO fundamental at $4.6 \mu\text{m}$ support this identification. In this letter we argue that the CO forms part of the supernova ejecta, and present a preliminary analysis of the first overtone band using Boltzmann population statistics. The model reproduces well the spectra obtained on days 192 and 255 and implies the presence of about 5×10^{-5} solar masses of CO by day 255. By day 284 the fit is less good, owing

to increasingly prominent emission from other species, possibly in combination with the breakdown of the Boltzmann distribution.

The spectra considered here and shown in Fig. 1a–e were obtained with a cooled grating spectrometer (FIGS) attached to the Anglo-Australian Telescope (AAT) as part of an extensive programme of infrared spectroscopy, to be described in more detail elsewhere (W.P.S.M. *et al.*, manuscript in preparation). The spectra were obtained 192, 255, 284 and 349 days after the explosion. The resolution, $\lambda/\Delta\lambda$, was 1,500 for all epochs, except for day 255, when it was 500. Absolute flux calibrations are believed to be reliable to 5–10% r.m.s. Errors are consistent with the scatter in the points of the spectra. Sharp features in the spectra should be interpreted with caution as the raw data in this spectral region contain many telluric features. In particular, we do not associate the sharp feature at $2.3 \mu\text{m}$ in the day-284 spectra with CO emission as it does not lie at a CO band head, but coincides with a known telluric feature. (Low-dispersion molecular spectra exhibit broad features known as bands which usually have a sharp edge at one end called a band head.)

Carbon monoxide has been observed before in stars^{5,6}, novae⁷, H II regions⁸ and galaxies⁹. The infrared spectra of SN1987A shown here provide convincing evidence that CO also occurs in association with young supernovae. The shape of the spectrum at $2.25\text{--}2.5 \mu\text{m}$ is consistent with emission originating from the $\Delta v = 2$ transitions of ground-state CO, and is similar to that observed and identified as CO in novae⁷. The discrete features observed correspond within the wavelength accuracy of the spectra to the band heads of $^{12}\text{C}^{16}\text{O}$. We therefore confirm the identification of these features as emission from the first overtone band of CO.

The potentially much stronger CO fundamental band is centred at $4.6 \mu\text{m}$. As early as day 117, a strong $4.6 \mu\text{m}$ excess was detected photometrically¹⁰; it has persisted to at least day 385 (ref. 11). To confirm an assignment to the CO fundamental, we obtained a suitable spectrum on day 192. The excess flux over the expected continuum suggested optically thick emission from CO. A similar excess was observed by Oliva *et al.*³ on about day 225, and on about day 262, observations from the Kuiper Airborne Observatory⁴ confirmed its identification with the $\Delta v = 1$ transitions of CO.

The $\Delta v = 3$ transitions might also be observed from the ground, but are expected to have about 1.4% the intensity of the first overtone¹². In the presence of strong features due to other species, we cannot identify them in our spectra.

To interpret the CO spectra, we must first ascertain its location. Specifically, does the CO lie in the progenitor wind, or has it formed in the supernova ejecta? Chevalier¹³ has argued that observed narrow ultraviolet lines originate from a swept-up shell of gas produced by the interaction of the winds from the blue supergiant progenitor and the previous red supergiant. He estimates the radius of the interaction zone to be $>2.4 \times 10^{17} \text{ cm}$. Any CO lying in the shell might be excited by the initial ultraviolet flash. But the wind velocity would be of the order of a few tens or hundreds of km s^{-1} , resulting in sharp features at the band heads. No such features are observed in our spectra, even at resolutions of 100 km s^{-1} . Moreover, the spectral shape should remain constant for the travel time of the ultraviolet flash through the shell. In contrast, the observed CO spectrum is changing continuously. Shock excitation by fast-moving ejecta of CO lying in the shell is also implausible: at the highest velocity seen, $\sim 40,000 \text{ km s}^{-1}$ (ref. 14), the travel time to the shell exceeds 2.5 years. We would in any case expect the shock to cause dissociation of CO. By contrast, the conditions prevailing in the ejecta at the epochs considered favour CO survival as a consequence of the very high molecular dissociation energy (11.09 eV; ref. 15). We conclude that the CO forms part of the supernova ejecta.

The observed intensity ratio of the fundamental to the first overtone band is an order of magnitude lower than expected in the optically thin case. But consideration of the energy levels

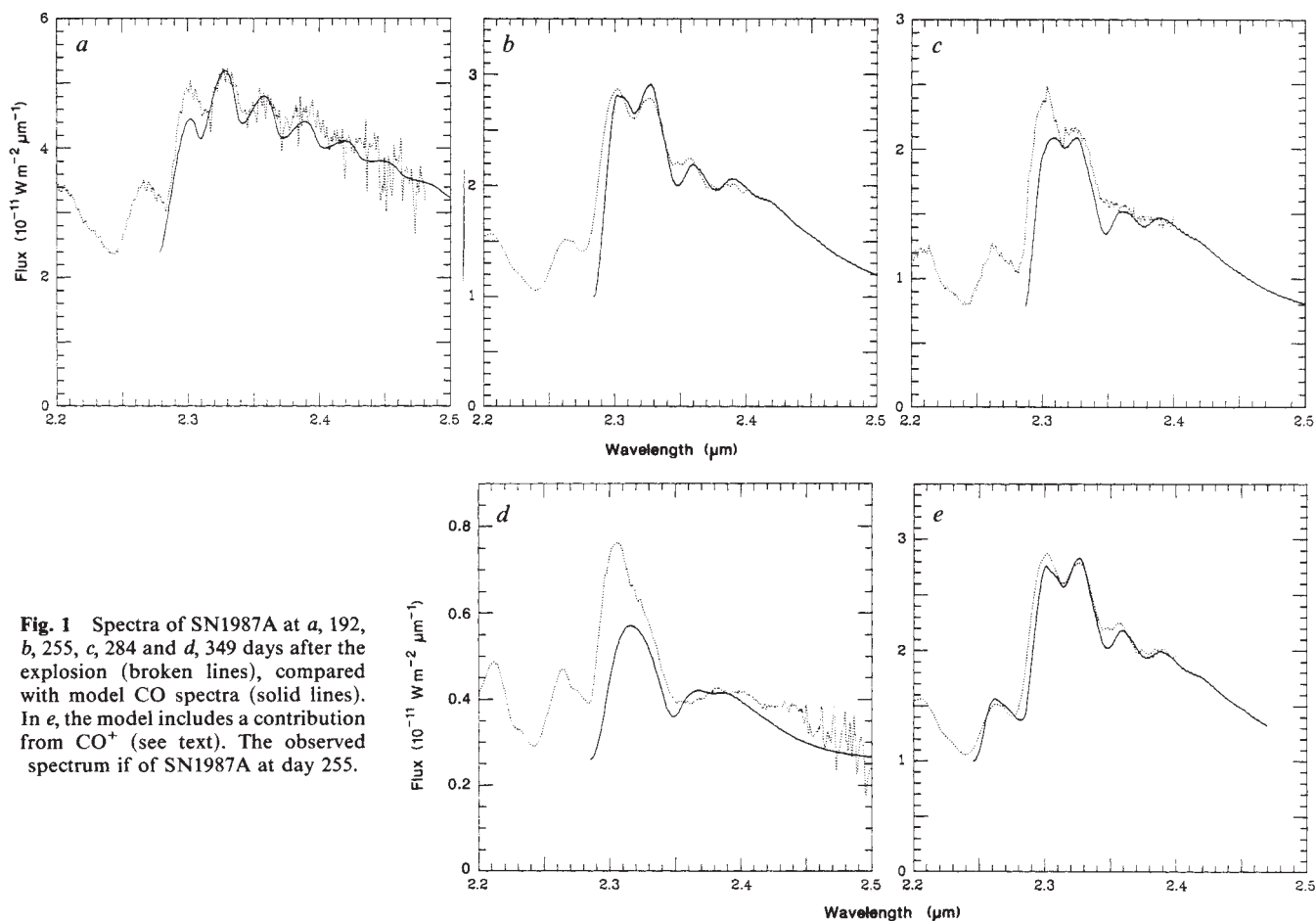


Fig. 1 Spectra of SN1987A at *a*, 192, *b*, 255, *c*, 284 and *d*, 349 days after the explosion (broken lines), compared with model CO spectra (solid lines). In *e*, the model includes a contribution from CO⁺ (see text). The observed spectrum of SN1987A at day 255.

and oscillator strengths shows that the nebula could indeed be optically thick to $\Delta v = 1$ photons while remaining optically thin in $\Delta v = 2$ (refs 16 and 3). Thus the observed ratio is understandable.

It has been noted¹⁷ that CO can dominate the opacity of the atmospheres of cool stars in the wavelength regions of the fundamental and overtones. The presence of CO in the ejecta of SN1987A could therefore be important to the structure and subsequent evolution of the supernova. To explore the conditions in the CO-emitting region, we have generated model spectra for the rotational-vibrational transitions of the ground states of $^{12}\text{C}^{16}\text{O}$ and $^{13}\text{C}^{16}\text{O}$.

The energy levels, E , were obtained from

$$E = \omega(v + \frac{1}{2}) - \omega x(v + \frac{1}{2})^2 + \omega y(v + \frac{1}{2})^3 + B_v J(J+1) - D_v J^2(J+1)^2$$

with values for ω , ωx , ωy , B_v and D_v from Young¹⁸ for $^{12}\text{C}^{16}\text{O}$, and from Benedict *et al.*¹⁹ for $^{13}\text{C}^{16}\text{O}$. The A-values, $A_{vv'}$, were obtained from

$$A_{vv'} = \frac{64\pi^4 w^3}{3hg_J} (TM)^2$$

where $TM = \langle vJ|M|v'J \pm 1 \rangle$ is the transition matrix for a transition between vibrational levels v and v' , and $g_J = 2J+1$. Chakerian and Tipping²⁰ calculated TM for the various isotopes of CO. The levels in our model were populated according to a Boltzmann distribution.

To derive the line profiles, we assumed the CO distribution to be spherical in extent, homologously expanding about the supernova centre of mass, and to have uniform density. The CO was also assumed to be optically thin in the first overtone emission. The resulting line profile is parabolic. In fact, isolated infrared emission lines of other species in SN1987A tend to be

Lorentzian in shape. Adoption of Lorentzian rather than parabolic profiles for CO does not significantly affect the resulting spectra. The free parameters are the mass of $^{12}\text{C}^{16}\text{O}$, the expansion velocity of the outer edge of the CO distribution and the temperature describing the populations of the CO levels. For each epoch the final model spectrum was obtained by convolving the line profile with the calculated CO spectrum, and adding the result to a continuum estimated by extrapolation from shorter wavelengths. The free parameters were fitted by visual comparison of the model with the data. This approach is adequate because the band heads are resolved in the data. The temperature is dictated by the ratios of the band heads of different v , the expansion velocity by the width of the band heads, and the mass of CO by the integrated intensity. We assumed the distance to the Large Magellanic Cloud (LMC) to be 50 kpc (ref. 21).

The derived values of these parameters are listed in Table 1, and the models and data are compared in Fig. 1*a-d*. For days 192 and 255, the fits are satisfactory at these epochs, vindicating

Table 1 CO model parameters

Epoch days	Continuum ($\text{W m}^{-2} \mu\text{m}^{-1}$)	Temperature (K)	Velocity (km s^{-1})	Mass ($M_\odot$)
192	2.38×10^{-11}	3,000	2,000	1.7×10^{-5}
255	1.00×10^{-11}	1,800	1,200	4.7×10^{-5}
284	0.71×10^{-11}	(1,600)	(1,200)	(1.2×10^{-4})
349	0.28×10^{-11}	(1,300)	(1,200)	(8×10^{-5})

For days 192–255 uncertainties in the derived temperature and velocity are of order 200 K and 100 km s^{-1} respectively. The uncertainty in the mass is $\sim 15\%$. For days 284 and 349, the parameters corresponding to the models displayed in Fig. 1*c* and *d* are shown in brackets as the models do not successfully reproduce the observed spectra at these epochs.

the use of a Boltzmann distribution. At these epochs the velocities are lower than observed in the hydrogen and helium lines, but comparable to those seen in oxygen, magnesium and iron lines. Thus some degree of stratification has been maintained during the explosion. The mass of CO increased over this 63-day period, but we cannot determine whether the increase resulted from CO forming in the ejecta or emerging through the shrinking photosphere, or both. The adequacy of the Boltzmann populations in the vibrational/rotational levels argues that the populating mechanism is predominantly collisional excitation, as suggested also by McGregor and Hyland². Other plausible populating mechanisms such as recombination, cascades from excited electronic states, and infrared pumping by the optically thick $\Delta v = 1$ transitions, would not result in a Boltzmann distribution¹⁶ and so cannot be significantly affecting the populations.

We also investigated the possible presence of $^{13}\text{C}^{16}\text{O}$ for days 192 and 255. The fit remains good, provided the ratio $N(^{12}\text{C})/N(^{13}\text{C}) > 10$, where N is the isotope number density. This lower limit on the $N(^{12}\text{C})/N(^{13}\text{C})$ ratio is similar to those found from novae⁷, and is consistent with the solar value of 80 (ref. 22).

In the later spectra (days 284 and 349), a less satisfactory match is obtained using Boltzmann populations. This is illustrated in Fig. 1c and d, where the Boltzmann model has been adjusted to match the P branch ($\sim 2.37\ \mu\text{m}$) intensity. Clearly, the model does not provide a simultaneous match to the R branch ($\sim 2.31\ \mu\text{m}$) feature. A reason for this could be deviation from a Boltzmann distribution in the vibrational level populations as a result of the increased significance of radiative excitation and de-excitation, or of recombination effects. At best, however, these effects can only provide a partial explanation for the failure of the Boltzmann model to reproduce the observed spectra. In the following, we argue that a further effect is the increasing prominence of emission from other species in this wavelength region.

For any specific rotational-vibrational transition, individual component pairs of the P and R branches (such as $\{R(0), P(2)\}$ or $\{R(1), P(3)\}$ and so forth) originate from the same upper level and have fixed intensity ratios. Moreover, the ratio is not strongly dependent on rotational quantum number. It follows that for a specific rotational-vibrational transition (for example, $v = 2 \rightarrow 0$), the ratio of the total flux in the R and P branches will be insensitive to the ambient temperature. We observe (Fig. 1d) that by day 349, the $v = 3 - 1$ band head has almost disappeared, presumably because cooling of the ejecta resulted in a negligibly small population in the $v = 3$ level. The first overtone emission is caused predominantly by radiation from the $v = 2$ level alone. In the corresponding model spectrum (Fig. 1d), the emission also originates mainly from the $v = 2$ level and the CO mass has been adjusted so that the model P branch matches the flux in the 2.35–2.40 μm region. The model R branch is about half the required intensity. We conclude that emission from some other species underlies the R branch. The alternative interpretation of absorption in the vicinity of the P branch seems unlikely.

From day 112 to day 349, the K-spectra also exhibit an emission feature at about 2.26 μm . It has been suggested that this feature could be due to CO^+ (S. J. Petuchowski, private communication). To investigate this proposal, we have generated an additional Boltzmann model spectrum with CO^+ included for day 255. The A-values for the vibrational-rotational transitions in the first overtone band of the ground state of CO^+ were obtained from Rosmus and Werner²³. The CO^+ was assumed to have the same temperature and velocity as was found in the neutral CO model fit for this epoch (Table 1). The mass of CO^+ was adjusted to provide a match to the 2.6 μm feature. We find that the addition of $0.62 \times 10^{-5} M_{\odot}$ of CO^+ improves the match of the model to the observations. This is illustrated in Fig. 1e (compare with Fig. 1b). The total carbon monoxide

mass ($M_{\text{CO}} + M_{\text{CO}^+}$) is $4.5 \times 10^{-5} M_{\odot}$, with a corresponding ionization fraction (CO^+/CO) of 0.16. We have also made a preliminary examination of the effect of adding CO^+ to the day-284 and day-349 model spectra. We find that whereas the addition of CO^+ can account for the 2.26 μm feature at these later epochs, the fit achieved at longer wavelengths is no better than is attained with the neutral CO model. There still appears to be emission in the 2.30 μm region whose origin is unidentified.

Finally, we point out that by day 255 the CO spectra indicate an ambient temperature below the condensation temperature of certain grain materials and so dust might be able to form²⁴. As yet, however, there is no evidence for dust in the ejecta.

We thank Mike Dopita, Eli Dwek, Bob Joseph and Sam Petuchowski for discussion.

Received 16 May; accepted 16 June 1988.

1. Catchpole, R. & Glass, I. *IAU Circ. No.* 4457 (1987).
2. McGregor, P. J. & Hyland, A. R. *IAU Circ. No.* 4468 (1987).
3. Oliva, E., Moorwood, A. F. M. & Danziger, I. J. *IAU Circ. No.* 4484 (1987).
4. Ames Research Center & Cohen, M. *IAU Circ. No.* 4500 (1987).
5. Thompson, R. I., Schnopper, H. W., Mitchell, R. I. & Johnson, H. L. *Astrophys. J.* **158**, L117–L122 (1969).
6. McGregor, P. J., Hyland, A. R. & Hillier, D. J. *Astrophys. J.* **324**, 1071–1098 (1988).
7. Ferland, G. J., Lambert, D. L., Netzer, H., Hall, D. N. B. & Ridgway, S. T. *Astrophys. J.* **227**, 489–496 (1979).
8. Osterbrock, D. E. *Astrophysics of Gaseous Nebulae*, 211 (Freeman, New York, 1974).
9. Young, J. S., Schloerb, F. P., Kenney, J. D. & Lord, S. D. *Astrophys. J.* **304**, 443–458 (1986).
10. Catchpole, R. N. *et al. Mon. Not. R. Astr. Soc.* **229**, 15P–25P (1987).
11. Whitelock, P. A. *et al. Mon. Not. R. Astr. Soc.* (in the press).
12. Bouanich, J. P. & Brodbeck, C. J. *Qu. Spectrosc. Radiat. Transf.* **14**, 1199–1208 (1974).
13. Chevalier, R. A. *Nature* **332**, 514–516 (1988).
14. Hanuschik, R. W. & Dachs, J. in *Proc. ESO Workshop on the SN1987A* (ed. Danziger, I. J.) 153–158 (European Southern Observatory, Garching, 1987).
15. Douglas, A. E. & Møller, C. K. *Can. J. Phys.* **33**, 125–132 (1955).
16. Scoville, N. Z., Krotov, R. & Wang, D. *Astrophys. J.* **240**, 929–939 (1980).
17. Alexander, D. R. & Johnson, H. R. *Astrophys. J.* **176**, 629–643 (1972).
18. Young, L. A. *Qu. Spectrosc. Radiat. Transf.* **8**, 693–716 (1968).
19. Benedict, W. S., Herman, R., More, G. E. & Silverman, S. *Astrophys. J.* **135**, 277–297 (1962).
20. Chakerian, C. Jr & Tipping, R. H. *J. molec. Spectrosc.* **99**, 431–449 (1983).
21. Rowan-Robinson, M. *The Cosmological Distance Ladder*, 147 (Freeman, New York, 1985).
22. Cameron, A. G. W. in *The Origin and Distribution of the Elements* (ed. Ahrens, L. H.) 125–143 (Pergamon, Oxford, 1968).
23. Rosmus, P. & Werner, H.-J. *Molec. Phys.* **47**, 661–672 (1982).
24. Dwek, E. *Astrophys. J.* **329**, 814–819 (1988).

Compact non-thermal radio emission from B-peculiar stars

R. B. Phillips* & J.-F. Lestrade†

* Massachusetts Institute of Technology, Haystack Observatory, Westford, Massachusetts 01886, USA

† Bureau des Longitudes, Paris 75014, France and Jet Propulsion Laboratory, Pasadena, California 91106, USA

Some stars hotter than 10,000 K show propensity for unusual surface abundances and excessive magnetic fields. These peculiar stars, Ap and Bp in the spectral nomenclature, show unusually prominent absorption lines of heavier elements. Rapid rotation and strong magnetic fields are revealed by line shapes and Zeeman splitting. We have used very-long-baseline interferometry (VLBI) to directly probe the source size and brightness temperature of weak radio emission recently discovered from two isolated Bp stars, σ Orionis E and HD37017. The emitting zone for each star is no more than 6 stellar diameters in extent, reflecting brightness temperatures of more than 10^9 K. Such high surface brightness resembles gyrosynchrotron radiation from mildly relativistic electrons trapped in the strong magnetic fields surrounding these stars^{1,2}. Compact radio radiation from these two stars presents new opportunities for probing the physical environments of early-type stars and for precise radio astrometry.

Bp stars, hotter and rarer than the Ap stars, include stars with abnormally high atmospheric abundance of He but apparently normal interior He abundances and masses. Main-sequence Bp He-rich stars include the two reported herein and were among

Table 1 VLBI measurements of the stars

Object	Scan start-stop (UT 8 March)	Raw fringe amplitude	SNR	Probability of false detection
σ Ori E	2036–2052	4.8×10^{-4}	7.7	5.3×10^{-9}
σ Ori E	2113–2129	5.4×10^{-4}	8.9	3.3×10^{-13}
HD37017	2149–2205	4.2×10^{-4}	6.4	4.9×10^{-5}
σ Ori E	2226–2242	5.5×10^{-4}	8.3	7.6×10^{-11}

Signal-to-noise ratio (SNR) and probability of false detection are based on number of independent samples recovered at playback; Rice-distributed statistics assumed.

several subclasses of magnetic Bp and Ap stars recently surveyed in the radio by Drake *et al.*² For 20 years their potent (up to 10 kilogauss) magnetic fields have encouraged radio surveys^{3–6} on the suspicion that some source of energetic particles might complete the recipe for radiation by the synchrotron process. None of these previous searches was successful, and set upper limits on flux density ranging from 5 to 50 mJy. Drake *et al.* detected 5 of 34 Ap and Bp stars selected for their strong magnetic fields by using the Very Large Array (VLA) near Socorro, New Mexico at 1.4, 5 and 15 GHz.

Radio emission from these stars was unresolved and weak, 1 to 3 mJy at 5 GHz. Radio spectra were flat or gently negative from 1.4 to 15 GHz, and varied by 10 to 30% over timescales of 1 to 10 days. Drake *et al.* persuasively argued the emission must be non-thermal, probably gyrosynchrotron from mildly relativistic electrons trapped temporarily in the intense magnetic fields around these stars. Linear sizes of their radio source models ranged from 1.5 to 10 stellar diameters ($R \approx 4R_{\odot}$).

At the distance of σ Ori E, 450 pc, the linear source size predicted by Drake *et al.* is equivalent to an apparent angular size of 0.001 arcsec or less. But the angular resolution of their successful VLA study was only about 0.1 arcsec, leading to a lower limit of 10^4 K for the brightness temperature. We were emboldened to organize a VLBI observation on an intercontinental baseline to achieve resolution better than 0.0005 arcsec, in order to measure a brightness temperature in the range expected for non-thermal emission ($T_b > 10^8$ K) or set an upper limit through failure to detect. Detecting radio sources of any type as faint as 2–3 mJy challenges the present capabilities of the VLBI technique. Only the most sensitive combinations of antennae and data acquisition system can be considered, and the sources must also be nearly unresolved. We computed that only the 100 m telescope of the Max-Planck Institut für Radioastronomie, near Effelsberg, FRG, and the entire 27-antennae phase-linked VLA could realistically serve to detect any of these stars at 5 GHz. As the noise level of VLBI or any radio observations is proportional to $(B\tau)^{-1/2}$, where B is the observed bandwidth and τ the integration time, we required the Haystack-NASA Mark III VLBI recording system ($B = 56$ MHz), and used experimental 13,000 ft extra-thin video tapes which allowed extension of τ beyond normal restrictions by 30%. The observations were made on 8 March 1988, under the auspices of the US VLBI Network and the two observatories. The recorded tapes were then shipped to Haystack for correlation on the Mark III VLBI processor.

Both σ Ori E and another Bp He-rich star, HD37017, were

detected in 16-minute integrations on the MPI-VLA baseline (Table 1). The raw VLBI amplitudes were calibrated using system temperatures and sensitivities measured at the antennae during the observations. Short recordings of strong extragalactic sources, two quasars, were included to help find the interferometer fringes in delay space and to check system performance. Some scheduled scans on these calibrator sources were missed, so we are unable to validate our *a priori* values for receiver and antenna performance for absolute accuracy better than $\pm 10\%$. Because these stars in Orion are south of the celestial equator, both telescopes were forced to observe at low elevations—though on opposite horizons. Resulting increases in system noise at these low elevations made the observations more difficult. The VLA collected imaging data in its normal way in parallel, which allowed simultaneous measurements of total flux density for σ Ori E and HD37017. We were concerned that a compact background source might be misidentified as one of the stars, and so mapped both fields using the VLA data. Intrinsic VLBI fringe-rate and delay windows on the sky were about 5 arcsec and 25 arcsec in right ascension and declination, respectively. No other sources are visible above the VLA map noise (~ 0.2 mJy) within a sky window five times that size. Residual VLBI fringe rates and delays for both stars and extragalactic calibrator sources are consistent with a small frequency standard offset ($\sim 3 \times 10^{-6} \mu\text{s s}^{-1}$) between the two observatories, but indicate no measurable position discrepancy from the VLA-derived radio star coordinates. Results for both stars are given in Table 2. Calibrated correlated flux densities and VLA total flux densities yield visibilities that are consistent with gaussian-shaped sources no more than 0.5 mas in size (full-width, half-maximum). When using such long coherent integrations as 16 min, frequency standard performance or atmospheric disturbances may become relevant to the coherent solution of the VLBI fringe-fitting process. We see no evidence for breaks in coherence, but any degradation would only tend to reduce fringe amplitudes and later bias the fitted source sizes to larger values. Our VLBI observations measure the source size only in the east-west direction (position angle 80°) on the sky. These VLBI detections contain no information about the fine structure of the radiospheres; the sky projection of the Effelsberg-VLA baseline did not change significantly in length or position angle during the observations.

Although the data indicate that we have barely resolved the radio sources, the measured visibilities in fact only establish useful upper bounds to source size, which can be stated as $0.3^{+0.2}_{-0.3}$ mas for both stars. Such a compact size is the first direct spatial evidence for a non-thermal origin to the radio emission of any early-type stars; the corresponding lower limit on the brightness temperature is 10^9 K.

σ Ori E is in a multiple (but detached) star system in the Orion complex of O and B stars. The distance estimate of 450 pc is based on accepted values for the absolute magnitudes of its normal neighbours. HD37017 is a spectroscopic binary of period 18.6 days⁷, but it is presumably part of the same larger Orion complex, and we adopt the same distance value. Our radio size limits correspond to linear sizes of 0.2 AU, or 3×10^{12} cm. The radio-emitting regions are no more than 6 stellar diameters across.

What is the significance of the compact, circumstellar radio radiation? σ Ori E has been the subject of optical study and

Table 2 Radio characteristics of the compact emission

Object	Spectral type	Right ascension (1950)	Declination (1950)	Correlated flux density (mJy)	Total flux density (mJy)	Linear size at 450 pc (AU)
σ Ori E	B2Vp	05 36 16.48	−02 37 18.6	3.1 ± 0.3	3.5 ± 0.3	≤ 0.2
HD37017	B1Vp	05 32 53.35	−04 31 31.9	2.0 ± 0.2	2.6 ± 0.2	≤ 0.2

Right ascension and declination are epoch 1950.0, radio position as measured at the time of observations (1988.2).

speculation for decades⁸. Strong, variable hydrogen- α emission was discovered by Walborn⁹ and spurred discoveries of other spectral anomalies of the same 1.2-day period^{10,11}. Both σ Ori E and HD37017 show weak stellar winds¹², and are believed to be oblique rapid rotators with strong, dipolar magnetic fields^{13,14}. Interior to some critical radius, the magnetic field of these models dominates the wind, forming what can correctly be termed a magnetosphere, with closed field lines. This hypothetical magnetic cavity is cospatial with the region of H α emission, and they may be intimately related. Outside the critical radius the wind dominates, forming a zone of open, spiral field lines like those of the sun. In the 'radiation belt' models of Drake *et al.*, the 5 GHz radio emission appears to come from a region 3 or 4 stellar radii out. Our VLBI-measured compact radiosphere is wholly consistent with such a picture. If a gyro-synchrotron process is assumed for the emission as Drake *et al.* suggest, the measured brightness temperatures and the analytical formulae of Dulk¹⁵ yield values of 150 gauss for magnetic field strength at the 5 GHz emission surface ($\tau = 1$), and a corresponding electron energy of 1 MeV if the electron-energy spectral index δ is 2. Such a 5 GHz emission surface would indeed be about 3 stellar radii above the stellar surface. We note that radiative lifetimes of the 5 GHz electrons are 0.5–1.0 days, implying surprisingly uniform replenishment to allow the 5 GHz emission to remain so constant. Synchrotron radiation cannot be completely ruled out on the basis of the measured physical characteristics, but a mechanism for producing the highly relativistic electrons needed is difficult to envision.

Although it is novel to find a hitherto unknown class of compact object, our detection of σ Ori E and HD37017 holds broader promise for the fields of stellar systems and atmospheres. Absolute scales of stellar brightness have been constructed like pyramids, founded on parallactic measurements of distance. Both σ Ori E and HD37017 should show annual trigonometric parallaxes of 2 mas if they are situated at the distances believed. Milliarcsecond-precision positions for RS CVn systems and the star Algol have been obtained with VLBI astrometry^{16,17}. Validation of the absolute magnitudes of main-sequence Bp stars can begin, using an independent astrometric method involving a totally different frequency regime and the stable grid of extragalactic sources.

The *Hipparcos* spacecraft will establish an astrometric grid of bright stars, which can be linked to the extragalactic VLBI reference frame through observations of the few radio-bright stellar systems. Radio centroids of symbiotic binaries may move about in some way, erratic or not, depending on orbital phase and activity. Solitary objects like σ Ori E are free of such drawbacks and make nearly ideal targets for the astrometric tying of radio and optical frames.

Stars such as σ Ori E present opportunities for astrometric companion searches. These two relatively distant Bp stars ($d = 450$ pc, $M_* \approx 7 M_\odot$) would show relatively small gravitational perturbations from dark companions in comparison with less massive stars ($M_* \leq 1 M_\odot$) closer to the solar neighbourhood. Even if Bp stars do not prove to be the candidates with which sensitive radio searches for extrasolar planets can best be made, we are optimistic that VLBI will detect the radiospheres of a variety of other stellar types in the coming decade. VLBI phase-referencing techniques already offer sufficient sub-milliarcsecond astrometric accuracy¹⁸ to make searches for massive dark companions feasible.

We thank the staff of the Max Planck Institut für Radioastronomie and of the National Radio Astronomy Observatory's Very Large Array, particularly Messrs Greschke, Hartley and Heatwole, and Haystack's Millson, Sellers and Sousa. S. A. Drake graciously shared his 1985 coordinates for both stars. This work was supported by grants from the National Science Foundation (R.B.P.) and the National Aeronautics and Space Administration (J.F.L.). The National Radio Astronomy Observatory's VLA is operated by Associated Universities, Inc.,

under contract with the NSF. Haystack Observatory is supported by the NSF through the Northeast Radio Observatory Corporation. The Jet Propulsion Laboratory of the California Institute of Technology is operated under contract with NASA.

Received 19 April; accepted 9 June 1988.

1. Lestrade, J.-F. *et al. Astrophys. J.* **282**, L23–L26 (1984).
2. Drake, S. A. *et al. Astrophys. J.* **322**, 902–908 (1987).
3. Trasco, J. D., Wood, H. J. & Roberts, M. S. *Astrophys. J.* **161**, L129–L131 (1970).
4. Kodaira, K. & Fomalont, E. B. *Astrophys. J.* **161**, 1169–1171 (1970).
5. Altenhoff, W., Pfänderer, J. & Weiss, W. in *IAU Colloquium 32, Physics of Ap Stars* (eds Weiss, W., Jenker, H. H. & Wood, H. J.) 497–501 (Reidel, Dordrecht, 1976).
6. Turner, K. C. in *Radio Stars* (eds Hjellming, R. M. & Gibson, D. M.) 283–286 (Reidel, Dordrecht, 1985).
7. Blaauw, A. & van Albada, T. S. *Astrophys. J.* **137**, 791–820 (1963).
8. Greenstein, J. L. & Wallerstein, G. *Astrophys. J.* **127**, 237–245 (1958).
9. Walborn, N. R. *Astrophys. J.* **191**, L95–L98 (1974).
10. Hesser, J. E., Walborn, N. R. & Ugarte, P. *Nature* **262**, 116–117 (1976).
11. Grootte, D. & Hunger, K. *Astr. Astrophys.* **56**, 129–133 (1977).
12. Barker, P. K. in *IAU Colloquium 87, Hydrogen Deficient Stars and Related Objects* (eds Hunger, K., Schöberner, D. & Kameswara, N.) (Reidel, Dordrecht, in the press).
13. Landstreet, J. D. & Borra, E. F. *Astrophys. J.* **224**, L5–L8 (1978).
14. Borra, E. F. & Landstreet, J. D. *Astrophys. J.* **228**, 809–816 (1979).
15. Dulk, G. A. *Rev. Astr. Astrophys.* **23**, 169–224 (1985).
16. Lestrade, J.-F., Preston, R. A., Mutel, R. L., Niell, A. E. & Phillips, R. B. in *Second FAST Thinkshop on the Processing of Scientific Data from the ESA Satellite HIPPARCOS* (ed. J. Kovalesky) 383–389 (CERGA, Grasse, France, 1987).
17. Lestrade, J.-F., Rogers, A. E. E., Niell, A. E. & Preston, R. A. in *IAU Symposium 129, The Impact of VLBI on Astrophysics and Geophysics* (eds Reid, M. J. & Moran, J. M.) (Reidel, Dordrecht, in the press).
18. Bartel, N., Herring, T. A., Ratner, M. I., Shapiro, I. I. & Corey, B. E. *Nature* **319**, 733–738 (1986).

High-magnetic-field critical currents in thin films of YBa₂Cu₃O₇

J. S. Satchell, R. G. Humphreys, N. G. Chew,
J. A. Edwards & M. J. Kane

Royal Signals and Radar Establishment, St Andrews Road,
Great Malvern WR14 3PS, UK

Critical current densities measured in bulk high- T_c superconductors are low, and fall dramatically in applied magnetic fields¹. This has raised doubts about the future use of high- T_c materials in superconducting machines and magnets. Thin films are known to perform much better in the absence of fields, but the effect of large magnetic fields on their critical currents has not been reported. Here we present measurements on high-quality polycrystalline thin films, which demonstrate that technologically useful current densities can be sustained at 77 K in high fields. The critical current density is strongly dependent on the direction of the applied field, reflecting the crystallographic anisotropy. Our results show that the material required to make high-field magnets will need to be crystallographically orientated, but not necessarily monocrystalline.

Films of Y–Ba–Cu–O were deposited by radio frequency magnetron sputtering from a metallic target, composed of a copper disk with inset pellets of yttrium and barium, in 15 m torr of argon/5% oxygen. The substrates were (001) SrTiO₃, their temperature during deposition was 400 °C and the deposition rate was 0.5 $\mu\text{m h}^{-1}$. The films were recrystallized under flowing oxygen at a maximum temperature of 850 °C. The film thickness was 0.5 μm and the transition temperature T_c (at zero resistance) was 85 K. At 77 K the critical current density was $5 \times 10^9 \text{ A m}^{-2}$. This is a very high value when compared to bulk material, but is somewhat less than the current state of the art for thin films^{2–4}, largely because of our lower T_c . The critical current density of this and similar, narrower tracks (see below) varied approximately linearly with $(T_c - T)$ between 83 and 20 K.

X-ray diffraction measurements indicate that the film used for this work was essentially single-phase YBa₂Cu₃O₇ and highly

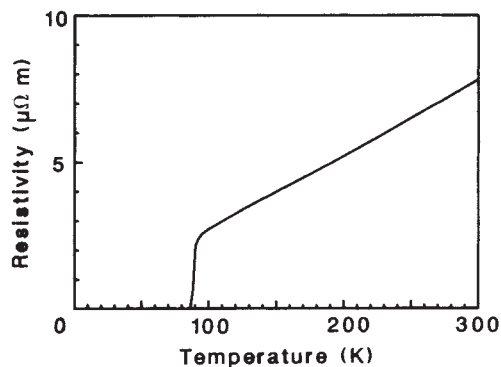


Fig. 1 Resistivity as a function of temperature for the track. The measurement current was $1 \mu\text{A}$, corresponding to a current density of $1.8 \times 10^4 \text{ A m}^{-2}$.

epitaxially oriented with respect to the substrate, with c -axes predominantly perpendicular to the substrate, but with $\sim 20\%$ a -axis orientation. Transmission electron microscopy on a similar film indicated that the epitaxially oriented grains were $\sim 1 \mu\text{m}$ in lateral size, with a proportion of smaller randomly oriented grains on top of the large ones.

The electrical properties were measured using the standard four-terminal technique on a narrow track, delineated by localized laser ablation using a continuous-wave 514.5-nm Ar^+ ion laser focused onto the sample through a $\times 40$ objective. The sample was translated under the stationary laser spot in a vacuum cell mounted on a computer-controlled x - y table. The effective width of the laser cut was estimated to be $3.3 \mu\text{m}$ by comparing the normal-state resistances of tracks of different nominal widths. The track used for the transport experiments was $11 \mu\text{m}$ wide and $300 \mu\text{m}$ long, and the layer was removed for $\sim 0.3 \text{ mm}$ either side of it; the contacts were 1-mm-diameter evaporated silver.

Figure 1 shows the resistivity of the track as a function of temperature. The normal-state resistivity shows 'metallic' behaviour, characteristic of conduction in the plane perpendicular to the c -axis³. The width of the transition is $\sim 2 \text{ K}$.

The supply used for the critical current measurements had a maximum output of 0.1 A , giving a measurement limit of $1.8 \times 10^{10} \text{ A m}^{-2}$. The critical current was taken to be the current which gave rise to a voltage of $1 \mu\text{V}$ across the sample. The voltage was measured one second after switching on the current. The sample was placed in the variable-temperature insert of a superconducting magnet and its temperature monitored with a carbon-glass resistance thermometer.

We report the critical current density for two different directions of applied field: (1) $\mathbf{B}$ perpendicular to the substrate and (2) $\mathbf{B}$ in the substrate plane, transverse to the track. Figure 2 shows the critical current as a function of field and temperature for orientations 1 and 2. Application of fields along the direction of conduction had a similar but smaller effect to orientation 2, indicating that the conduction is probably percolative, with significant local transverse current flow. The data have been slightly corrected to allow for temperature variations during the field sweep.

The critical current density (J_c) fell to zero in a field of 6.5 T for orientation 1 ($\mathbf{B} \parallel c$) at the highest temperature at which we took data (78 K). This is close to the upper critical field (B_{c2}) of 5 T , predicted from our measured T_c and $dB_{c2}/dT(\mathbf{B} \parallel c) = -0.71 \text{ T K}^{-1}$ from ref. 5. In orientation 2 ($dB_{c2}/dT = -2.3 \text{ T K}^{-1}$) the upper critical field is still large ($\sim 16 \text{ T}$) at 78 K ; the measured critical current density was $3 \times 10^8 \text{ A m}^{-2}$ at 6.5 T . This is at the level required for many applications, and would doubtless be larger in material with optimized T_c .

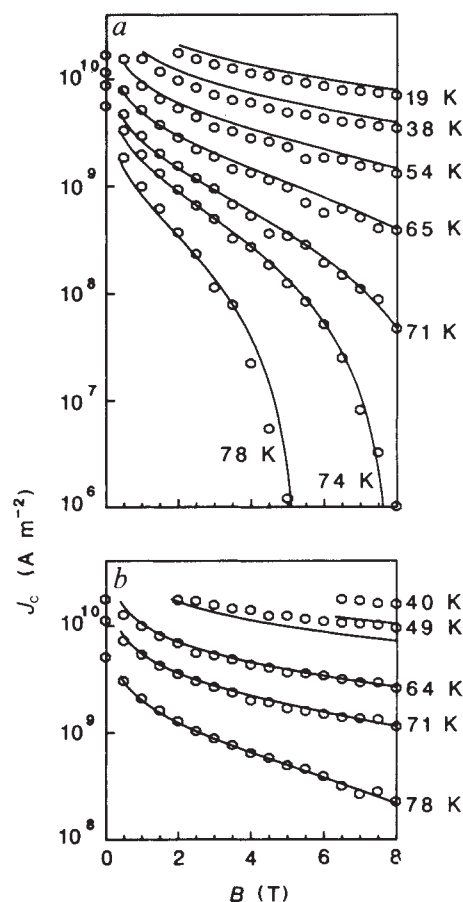


Fig. 2 Field dependence of critical current for orientations 1(a) and 2(b) at the temperatures indicated. The critical currents given by equation (1) are shown by the smooth curves, with $\alpha = 1.0 \times 10^8 \text{ H}^{-1}$ (a) and $3.3 \times 10^7 \text{ H}^{-1}$ (b).

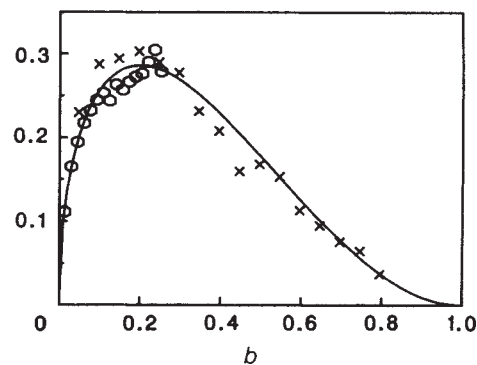


Fig. 3 Dimensionless form of the flux pinning force ($f = F_p / (\alpha B_{c2}^2)$) plotted as a function of reduced field, b , showing that the data for both orientations lie on the curve given by equation (1). Only data at 71 K are shown for clarity; $\times$, orientation 1, $\circ$, orientation 2.

The critical current of a type II superconductor at high fields ($> B_{c1}$) is controlled by the extent to which the flux penetrating the material is free to move. In conventional superconductors, the flux pinning force density, $F_p = \mathbf{J}_c \times \mathbf{B}$, can often be represented^{6,7} in terms of the products of powers of the reduced field, $b \equiv B/B_{c2}$, $(1-b)$ and B_{c2} . Our results at high temperatures

($T \geq 55$ K) can be well represented by

$$F_p = \alpha B_{c2}^2 b^{1/2} (1-b)^2 \quad (1)$$

where α is an empirical constant. Equation (1) is represented by the solid lines in Fig. 2, and in more detail in Fig. 3. This form of b -dependence is similar to the pinning force found in superconductors with equiaxed grains and pinning at the grain boundaries^{6,7}, but the dependence on $B_{c2}(T)$ is slightly weaker. The data at lower temperatures cover a restricted range of reduced fields, but clearly deviate from this form, showing increased anisotropy. We are unable to distinguish any effects of surface pinning in orientation 2 from the large crystallographic anisotropy. In orientation 1, edge pinning is expected to be relatively unimportant. This is confirmed by the observation that a wider track ($40 \mu\text{m}$) also gave very similar pinning forces.

All of the data presented here were obtained for increasing field. Subsequent measurements on another track with lower J_c ($1.6 \times 10^9 \text{ A m}^{-2}$ at 77 K) showed the same flux pinning law (equation (1)), but with smaller values of α . Hysteresis between increasing and decreasing field sweeps was minimal.

We thank A. G. Cullis for the transmission electron microscopy and O. D. Dosser for EDX composition measurements.

Received 17 June; accepted 6 July 1988.

1. Ekin, J. W. *J. appl. Phys.* **62**, 4821–4828 (1987).
2. Koch, R. H. & Chaudhari, P. *Nature* **332**, 682–683 (1988).
3. Enomoto, Y., Murakami, T., Suzuki, M. & Morikawa, K. *Jap. J. appl. Phys.* **26**, L1248–L1250 (1987).
4. Mankiewicz, P. M. *et al. Appl. Phys. Lett.* **51**, 1753–1755 (1987).
5. Worthington, T. K., Gallagher, W. J. & Dinger, T. R. *Phys. Rev. Lett.* **59**, 1160–1163 (1987).
6. Dew-Hughes, D. *Phil. Mag.* **B55**, 459–479 (1987).
7. Campbell, A. M. & Evetts, J. E. *Critical Currents in Superconductors* (Taylor and Francis, London, 1972).

Climatic impact of ice-age aerosols

L. D. Danny Harvey

Department of Geography, University of Toronto,
100 St George Street, Toronto M5S 1A1, Canada

Data from Antarctic^{1–3}, Greenland^{4,5} and Devon Island⁶ ice cores indicate that large increases in the atmospheric aerosol loading occurred during the Last Glacial Maximum (LCM) of about 18,000 yr BP. The aerosol content of the atmosphere is important in the present climate, causing the global mean temperature to be about 2–3 °C cooler than it would be in the absence of aerosols⁷. Here I use an energy balance climate model to show that plausible increases in the atmospheric aerosol optical depth during the LGM could have caused a further global mean cooling of 2–3 °C, thereby making a significant contribution to the climatic cooling of the LGM.

The energy balance climate model (EBCM) used here⁸ has surface-air, land-sea and latitudinal resolution, and is forced with seasonally varying but diurnally averaged insolation. It is coupled to a sea ice submodel⁹. The model sensitivity to solar constant variations of 2% is analysed and compared with other models elsewhere¹⁰; it is sufficient for our purposes to note that the model used here has a global mean atmospheric temperature response of –3.7 °C for a 2% solar constant decrease, and 3.4 °C for a 2% solar constant increase. The external forcing change ΔQ for these experiments is $\pm 4.9 \text{ W m}^{-2}$, giving a sensitivity parameter $\lambda = \Delta Q / \Delta T$ of 1.3–1.4, which is comparable to that obtained from many other models^{11,12}.

In calculating solar fluxes, separate calculations are performed in the visible and near-infrared parts of the spectrum and for cases of clear and cloudy sky. Aerosols are incorporated

Table 1 Optical depths for marine and continental aerosols over land and sea for the present climate

Latitude	Ocean		Land	
	Marine	Continental	Marine	Continental
85° N	0.011	0.05	0.011	0.05
75° N	0.013	0.05	0.013	0.09
65° N	0.023	0.05	0.015	0.14
55° N	0.029	0.06	0.014	0.19
45° N	0.036	0.05	0.018	0.23
35° N	0.036	0.04	0.018	0.33
25° N	0.040	0.05	0.020	0.38
15° N	0.040	0.06	0.025	0.38
5° N	0.041	0.07	0.025	0.32
5° S	0.039	0.06	0.028	0.23
15° S	0.036	0.05	0.031	0.24
25° S	0.040	0.05	0.034	0.27
35° S	0.042	0.04	0.038	0.16
45° S	0.042	0.03	0.038	0.10
55° S	0.039	0.02	0.033	0.05
65° S	0.012	0.01	0.012	0.01
75° S	0.012	0.01	0.012	0.01
85° S	0.012	0.01	0.012	0.01

using the parameterization in ref. 7, except that the delta-Eddington equations are used instead of the two-stream approximation. Aerosols are restricted to the part of the solar spectrum from 0.3 to 0.9 μm in wavelength and are assumed to lie completely under clouds for the cloudy-sky calculations, and to completely underlie Rayleigh scatterers for the clear-sky calculations. The direct radiative forcing of aerosols is dependent on multiple reflections involving the surface, clouds and Rayleigh scatterers, all of which are included in the model used here.

Zonally averaged aerosol optical depths at 0.55 μm are separately specified for marine and continental aerosol types over both land and sea (Table 1) and were obtained from ref. 13. The resultant area-weighted global mean total aerosol optical depth τ is 0.134, which agrees with the estimate of $0.100 < \tau < 0.175$ obtained previously¹⁴. An asymmetry factor of 0.65 is used for both aerosol types, and single-scattering albedos ω of 0.986 and 0.940 are used for marine¹³ and continental⁷ aerosols respectively. Other workers¹³ used an ω of 0.87 for continental aerosols, which is somewhat more absorbing than I use here and is more appropriate for Saharan dust¹⁵. Sensitivity tests here use both single-scattering albedos for continental aerosols. The two aerosol types are combined into a single aerosol layer by adding the separate optical depths and using the optical-depth weighted values of the single-scattering albedos and asymmetry factors.

The climatic response to an increase in aerosols depends on (1) the magnitude of the change in radiative heating at the instant of the aerosol increase, but before any climatic response has occurred and (2) the strengths of the various positive and negative climatic feedback processes. The first factor is the change in external forcing, ΔQ , and the second factor determines how large the climatic response will be for a given ΔQ . The relative importance of changes in marine and continental aerosol optical depths to the climate of the LGM can be estimated by comparing the corresponding ΔQ s. Table 2 gives mean annual globally averaged ΔQ s when marine and continental aerosol optical depths are separately and uniformly increased, both alone and in the presence of the other aerosol type. For continental aerosols, results are shown for single-scattering albedos of 0.87 and 0.94. The direct radiative effect of increases in both aerosol types is one of cooling, with continental aerosols being only ~60% as effective as marine aerosols for $\omega = 0.87$, and 80% as effective for $\omega = 0.94$. The presence of marine aerosols with optical depth 0.1 reduces the effect of changes in continental aerosols by about 10%, whereas the presence of continental aerosols reduces the effect of changes in marine aerosol optical depth by 6–18% for continental optical depths of 0.1–0.3. Figure

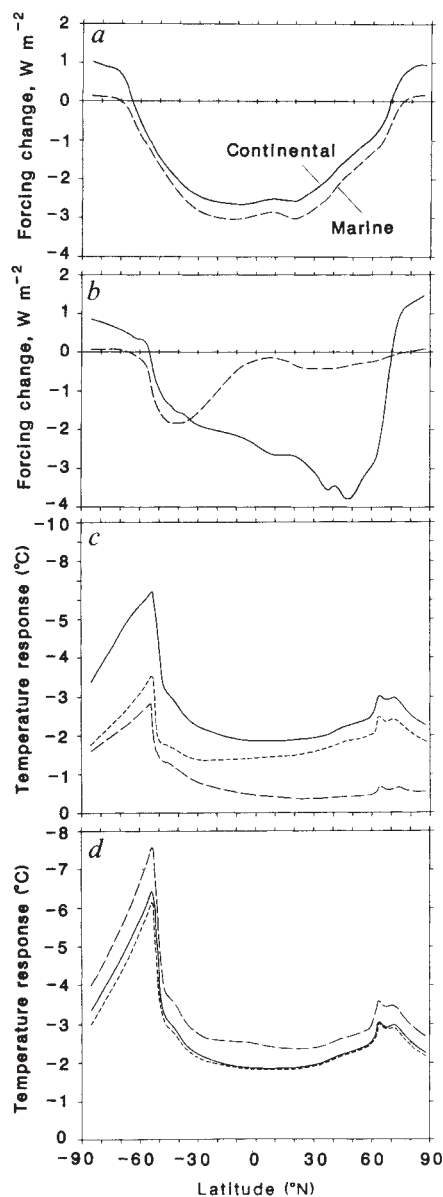


Fig. 1 *a*, Zonally averaged mean annual forcing change in going from an aerosol-free atmosphere to continental aerosols only with optical depth 0.1, and to marine aerosols only with optical depth 0.1. *b*, Zonally averaged mean annual forcing change resulting from the separate base case Last Glacial Maximum increases in marine and continental aerosols. *c*, Zonally averaged mean annual atmospheric temperature responses for the base case Last Glacial Maximum increase in marine aerosols (long dashes), continental aerosols (short dashes), and marine and continental aerosols together. *d*, Zonally averaged mean annual atmospheric temperature response for the base case Last Glacial Maximum aerosol increase (solid line) and when the aerosol optical depth is increased by 0.81 at 60–90° N and 60–90° S only (short dashed line) and at 25° S–25° N only (long dashed line).

1*a* shows the latitudinal variation of ΔQ when continental and marine aerosols with optical depth 0.1 are added to an aerosol-free atmosphere. As has also been previously found^{7,16}, the direct radiative forcing is one of heating at high latitudes, where high surface albedos occur, and one of cooling elsewhere.

Data from Antarctic ice cores indicate that the aerosol concentration in the ice was 4–6 times and 10–30 times greater during the LGM than during the Holocene for marine aerosols and continental aerosols respectively². Assuming that 40% of marine and continental aerosol deposited during the LGM

Table 2 Change in mean annual globally averaged radiative forcing, ΔQ , for globally uniform changes in aerosol optical depth, τ

Case	Forcing change (W m^{-2})
$\tau_{\text{cont}} = 0$	
τ_{marine} 0.0 to 0.1	-2.44
0.1 to 0.2	-2.29
0.2 to 0.3	-2.15
$\tau_{\text{marine}} = 0$, $\omega_{\text{cont}} = 0.94$	
τ_{cont} 0.0 to 0.1	-2.03
0.1 to 0.2	-1.87
0.2 to 0.3	-1.73
$\tau_{\text{marine}} = 0$, $\omega_{\text{cont}} = 0.87$	
τ_{cont} 0.0 to 0.1	-1.45
0.1 to 0.2	-1.30
0.2 to 0.3	-1.17
$\tau_{\text{marine}} = 0.1$, $\omega_{\text{cont}} = 0.94$	
τ_{cont} 0.0 to 0.1	-1.87
$\tau_{\text{marine}} = 0.1$, $\omega_{\text{cont}} = 0.94$	
$\tau_{\text{cont}} = 0.1$	-2.28
$\tau_{\text{cont}} = 0.3$	-2.00

occurred as wet deposition and 60% as dry deposition, and taking the snowfall rate during the LGM to be half that of the Holocene^{17,18}, the atmospheric marine and continental aerosol loadings were respectively ~3–4.6 and 8–25 times greater during the LGM over Antarctica. The continental aerosols are of desert origin and presumably originated from expanded low to mid-latitude deserts, where it has been speculated³ that the aerosol optical depth may have been 2–4 times the present optical depth, assuming no change in the effectiveness of atmospheric aerosol transport to high latitudes. The marine aerosols over Antarctica, on the other hand, were probably derived from the nearby southern Ocean^{1,2}. Note that an increase in marine aerosol deposition occurred in spite of the probably greater distance from the sea ice margin to the Antarctic ice cap during the LGM indicated by the CLIMAP reconstruction¹⁹; this strongly implies that a significant increase in atmospheric marine aerosol source strength and loading also occurred over the southern Ocean. In the Northern Hemisphere, at Camp Century, Greenland, the concentration of insoluble particles is 12 times greater for the LGM than during the Holocene⁴, whereas in profiles from the Devon Island ice cap the insoluble aerosol concentration is 20–30 times the Holocene value during the LGM (ref. 6). For marine aerosols, the concentration during the LGM is about 3 times higher than during the Holocene at Camp Century⁵. On the basis of this evidence, the continental aerosol optical depth is assumed to increase by a factor of 2 between 35° S and 35° N, with a linear increase in the multiplication factor to a value of 6 at 90° N and a value of 8 at 90° S. Marine aerosol optical depth is assumed to increase by a factor of 2.0 poleward of 60° N, and by a factor of 3.8 poleward of 60° S, with a linear decrease of the multiplication factor to 1.0 at the equator. The same multiplication factors are applied over land and sea, and the resultant optical depths are such that the sign of the gross meridional gradient of aerosol optical depth is not changed between the present and the LGM.

Table 3 gives the mean annual globally averaged direct forcing and atmospheric temperature response for various experiments. For a doubling of present aerosol amounts everywhere, the temperature responses are -1.37 °C and -1.85 °C, using continental aerosol single scattering albedos of 0.87 and 0.94 respectively. The climate model response depends in part on the strength of ice and snow feedbacks, which will be stronger the larger the initial ice and snow areas and hence the colder the starting climate. Polar ice cap evidence indicates that large aerosol increases occurred only during the last phase of the last glaciation^{1–3}, and hence after some climatic cooling and

Table 3 Mean annual globally averaged atmospheric temperature response (ΔT_a) and radiative forcing change (ΔQ) for ice-age aerosol scenarios described in the text

Experiment	ΔT_a (°C)	ΔQ (W m ⁻²)
Present solar constant		
2 × present aerosols, $\omega_{\text{cont}} = 0.87$	-1.37	-1.89
2 × present aerosols, $\omega_{\text{cont}} = 0.94$	-1.85	-2.53
0.98 × present solar constant, $\omega_{\text{cont}} = 0.94$		
2 × present aerosols,	-2.26	-2.42
LGM marine aerosol increase	-0.77	-0.59
LGM continental aerosol increase	-1.72	-2.04
LGM base case aerosol increase	-2.52	-2.59
LGM base case + uncertainty in polar regions	-2.44	-2.55
LGM base case + uncertainty in low latitudes	-3.10	-3.31

The first three experiments shown involve an aerosol increase by a globally uniform factor of 2, whereas the multiplication factor varies regionally for all the other experiments, as described in the text. In all except the first two experiments, a reduced solar constant is used to give greater initial ice and snow areas before the aerosol increase.

extension of ice and snow had already occurred. The effect of greater initial ice and snow area on model sensitivity to an aerosol increase can be tested by first reducing the solar constant by 2%, then applying an aerosol increase. The temperature response due to an aerosol doubling is -2.26 °C for this case, using a continental single-scattering albedo ω_{cont} of 0.94 (Table 3).

In the remaining experiments, the LGM aerosol increase scenarios outlined above are used, the solar constant is first reduced by 2%, and $\omega_{\text{cont}} = 0.94$. Figure 1b shows the mean annual, zonally averaged ΔQ for marine and continental aerosols, and Fig. 1c shows the mean annual zonally averaged atmospheric temperature response when marine aerosols are increased alone, or continental aerosols are increased alone, and when marine and continental aerosols are increased together. In each case the response is one of cooling at all latitudes, in spite of the fact that the direct radiative forcing change at high latitudes is one of heating. The overall model response is thus largely governed by the global mean change in radiative forcing and positive feedbacks involving snow and sea ice, which are responsible for the greater response at high latitudes. The behaviour in response to aerosol increases is similar to that obtained from other models^{7,16}. For the base case situation, the temperature response is -2.5 °C or approximately half the global mean sea surface and sea ice surface cooling of 4–6 °C, as inferred for the LGM from CLIMAP and climate model results^{20,21}. Thus, possible aerosol increases during the LGM could have made a significant and largely overlooked contribution to the climatic cooling of the LGM.

In Fig. 1d, the relative importance of uncertainties in the aerosol loading at high and low latitudes is investigated. The mean annual zonally averaged atmospheric temperature response is shown for the base case LGM scenario (see Fig. 1c). The response when the continental aerosol loading is further increased between 60°–90° N and 60°–90° S by 6 times the initial aerosol loading over Antarctica (or 0.075) is also shown, together with the response when the same absolute increase is added to the aerosol optical depth in the 25° S–25° N zone (where it represents a much smaller relative uncertainty). Two interesting points emerge: (1) uncertainties in the aerosol optical depth at high and low latitudes have opposite effects on the temperature response at all latitudes, and (2) uncertainties at low latitudes (where we have no information on aerosol loadings) have a much greater effect on the global temperature response than the same absolute uncertainty at high latitudes. If a larger optical

depth at polar latitudes implies a larger optical depth at low latitudes, then we obtain the somewhat paradoxical result that the larger the direct warming effect of aerosols in the only regions where we have evidence of how they changed, then the greater is the temperature decrease, both in these regions and globally.

For the base case LGM aerosol scenario, the oceanic mixed-layer temperature cools by about 1.7 °C between 10° S and 10° N. According to the CLIMAP reconstruction¹⁹, sea surface cooling at these latitudes was only 1.5 °C, although there is evidence that the CLIMAP cooling may be 2 °C too small²¹. Nevertheless, the magnitude of inferred tropical sea surface temperature changes during the LGM could serve as a constraint on the aerosol optical depth increase at low latitudes, particularly as aerosol effects would operate in addition to such effects as reduced CO₂ and altered high-latitude boundary conditions.

Another possible route in removing the uncertainties in aerosol optical depths during the LGM is through simulation of the entrainment, transport and deposition of aerosols with general circulation models²². One implication of the results presented here is that continental aerosol optical depths in low-latitude source areas are particularly important to the global climatic response, suggesting that particular attention will have to be given to the entrainment processes and reconstruction of the past extent of arid regions.

Time series from Antarctic ice cores indicate that aerosol increases occurred after significant cooling had already occurred^{1–3}; hence, aerosol increases probably served as a positive feedback mechanism, amplifying the initial temperature decrease, rather than serving as a driving mechanism for glacial-interglacial oscillations. This is in contrast to atmospheric CO₂ variations, which appear to have led ice-volume variations and therefore may be part of the driving mechanism for glacial-interglacial oscillations^{23,24}. Also note that increases of atmospheric aerosols in low latitudes serve as a potential mechanism for inducing synchronous climatic change in both hemispheres; recent hypotheses focus on CO₂ variations to the complete exclusion of aerosol increases as a mechanism for linking the two hemispheres^{25,26}.

This work was supported by the Natural Sciences and Engineering Research Council of Canada using the Ontario Centre for Large Scale Computation Cray XMP. The initial phase of this research was performed at the Laboratoire de Glaciologie (Grenoble) and the Laboratoire de Meteorologie Dynamique (Paris), and the hospitality and use of facilities at these two institutes is appreciated.

Received 8 February; accepted 16 June 1988.

- De Angelis, M., Barkov, N. I. & Petrov, V. N. *Nature* **325**, 318–321 (1987).
- Petit, J. R., Briat, M. & Royer, A. *Nature* **293**, 391–394 (1981).
- Royer, A., De Angelis, M. & Petit, J. R. *Climat. Change* **5**, 381–412 (1983).
- Thompson, L. G. & Mosley-Thompson, E. *Science* **212**, 812–815 (1981).
- Cragin, J. H., Herron, M. M., Langway, C. C. Jr & Klouda, G. in *Polar Oceans* (ed. Dunbar, M. J.) 617–631 (Arctic Institute of North America, Calgary, 1977).
- Fisher, D. A. *Quat. Res.* **11**, 299–305 (1979).
- Coakley, J. A., Cess, R. D. & Yurevich, F. B. *J. Atmos. Sci.* **40**, 116–138 (1983).
- Harvey, L. D. D. *J. Clim.* (in the press).
- Harvey, L. D. D. *J. Clim.* (in the press).
- Harvey, L. D. D. *Clim. Change* (in the press).
- Dickinson, R. E., Meehl, G. A. & Washington, W. M. *Climat. Change* **10**, 241–248 (1987).
- Schlesinger, M. E. & Mitchell, J. F. B. *Rev. Geophys.* **25**, 760–798 (1987).
- Tanre, D., Geleyn, J. F. & Slingo, J. in *Workshop on aerosols and their climatic effects* (Williamsburg, Virginia, 1983).
- Toon, O. B. & Pollack, J. P. *J. appl. Met.* **15**, 225–246 (1976).
- Carlson, T. N. & Benjamin, S. G. *J. Atmos. Sci.* **37**, 193–213 (1980).
- Potter, G. L. & Cess, R. D. *J. geophys. Res.* **89**, 9521–9526 (1984).
- Yiou, F., Raisbeck, G. M., Bourles, D., Lorius, C. & Barkov, N. I. *Nature* **316**, 616–617 (1985).
- Jouzel, J. et al. *Nature* **329**, 403–408 (1987).
- CLIMAP Project Members *Seasonal Reconstruction of the Earth's Surface at the Last Glacial Maximum* (Geol. Soc. Amer. Map and Chart Series MC-36, 1981).
- Kutzbach, J. E. & Guetter, P. J. *J. Atmos. Sci.* **43**, 1726–1759 (1986).
- Rind, D. & Petet, D. *Quat. Res.* **24**, 1–22 (1985).
- Joussau, S., Rasool, & Sadourny, R. *Ann. Glaciol.* **5**, 204–207 (1983).
- Shackleton, N. J. & Pisias, N. G. in *The Carbon Cycle and Atmospheric CO₂ Natural Variations Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 303–317 (AGU, Washington, 1985).
- Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
- Manabe, S. & Broccoli, A. J. *J. Atmos. Sci.* **42**, 2643–2651 (1985).
- Denton, G. H., Hughes, T. J. & Karlen, W. *Quat. Res.* **26**, 3–26 (1986).

SIMS depth profiles of weathered plagioclase and processes affecting dissolved Al and Si in some acidic soil solutions

H. W. Nesbitt & I. J. Muir*

Department of Geology and * Department of Chemistry, University of Western Ontario, London, Ontario N6A 5B7, Canada

Many natural catchments are now under threat of acidification from natural and anthropogenic sources. Acids introduced to the catchments are neutralized by reaction with soil constituents (including clay minerals and feldspars) and a better understanding of the mechanisms of neutralization is required to evaluate effects of acid deposition in sensitive catchments. The response of feldspars to these acids has been generally studied by monitoring the composition of the weathering solutions. The dissolution of albite and perthite is initially incongruent, with Na- and K-depleted residues produced at the expense of the feldspars¹⁻⁴. Such residues can be detected by secondary ion mass spectrometry using the specimen isolation technique provided the residues are greater than ~40 Å thick. Here we report secondary ion mass spectrometric studies of naturally weathered feldspar surfaces. The acidic soil waters of many catchments have high contents of total dissolved aluminium⁵⁻¹⁰, including some from the Plastic Lake Catchment in Ontario, and we consider the effects of feldspar dissolution on Al contents of acidic soil waters.

Secondary ion mass spectrometry (SIMS) has been applied to many geological problems¹¹⁻¹⁶, but depth profiling of minerals using SIMS is limited¹⁷⁻¹⁹. It is used here to document the near-surface composition of naturally weathered plagioclase. Samples were obtained from a 10,000-year-old thin till deposited in the Plastic Lake Catchment. Stones from the till were cut to produce specimens (2×2×1 cm) that contain one naturally weathered face, then rinsed with tap water for ten seconds and air-dried. Plagioclase is of oligoclase composition (from electron microprobe analyses of the freshly cut surfaces). Sample surfaces were not cleaned to avoid possible destruction of delicate structures, textures and films related to feldspar alteration. Materials unrelated to feldspar weathering could be attached to the unwashed surfaces, but we decided it was better to identify these by textural, structural and compositional observations, rather than risk removal of any delicate products of feldspar weathering.

A CAMECA ims 3f secondary ion microscope was used in specimen isolation mode^{20,21} to remove molecular ions from the spectra. Most importantly, no conductive surface coating is required to analyse the non-conducting feldspar samples in the specimen isolation mode; there is no contamination of the plagioclase surfaces. A primary beam of mass filtered ¹⁶O⁺ ions at a net 15.0–16.0 kV and ~110 nA was rastered over a 250×250 μm area. To reduce crater edge effects during depth profiling, ions were sampled from a 60-μm diameter area located well within the rastered zone by using an aperture at the entrance to the spectrometer. A highly polished face of an oligoclase crystal was sputtered for 20 min and the depth of the crater measured using a Sloan Dektak profilometer. Three determinations gave a sputtering rate of ~55 Å per min (beam rastered over an area of 250×250 μm at a current of 110 nA). The irregular topography of the naturally weathered grains may produce sputtering rates differing by as much as 50 per cent from the measured rate, hence the thickness of the residual layers (discussed below) could differ by up to 50 per cent from our estimate.

A SIMS depth profile of an unweathered plagioclase grain exposed on a freshly cut surface is illustrated in Fig. 1a. Data for the main elements (Na, Al, Si, K, Ca) are included (Fig.

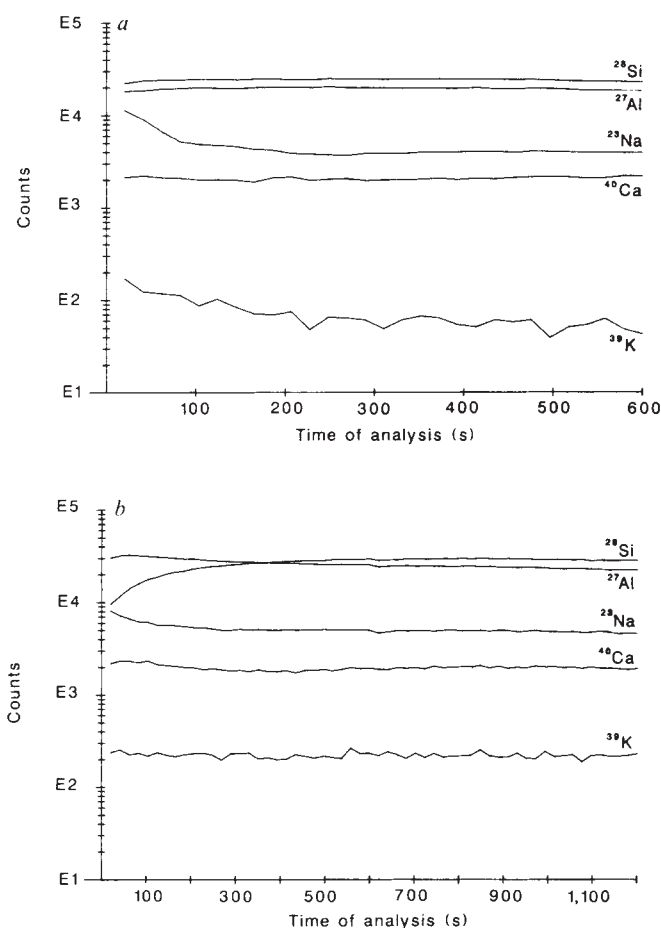


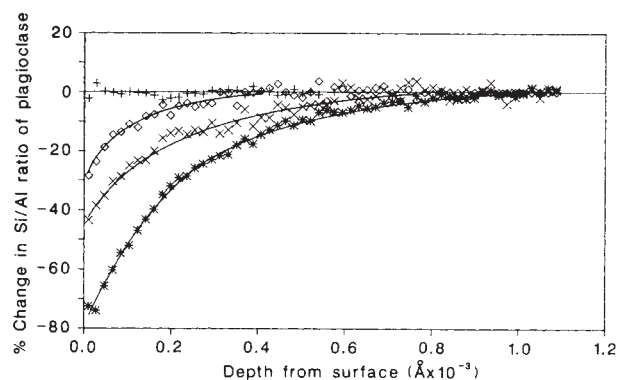
Fig. 1 a, SIMS depth profile from freshly cut surface to approximately 500 Å depth, of an unweathered plagioclase grain (Plastic Lake Catchment, Ontario, Canada). Observed counts (nominal number of secondary ions per second) are plotted against time of analysis (depth of sampling) for isotopes ²³Na, ²⁷Al, ²⁸Si, ³⁹K and ⁴⁰Ca. b, SIMS depth profile, from weathered surface to approximately 1,100 Å depth, of naturally weathered plagioclase grain no. 3 (plastic Lake Catchment, Ontario, Canada). Observed counts are plotted against time of analysis (depth of sampling) for isotopes ²³Na, ²⁷Al, ²⁸Si, ³⁹K and ⁴⁰Ca. Observed counts = nominal number of secondary ions measured per second.

1a). The ²⁸Si, ²⁷Al and ⁴⁰Ca profiles do not change appreciably over the time of the analysis (600 s). The ²³Na profile rises near the surface, but the same trend was observed during analysis of all the fresh plagioclase grains we studied (total, 39). There are no sympathetic changes to the Al, Si and Ca profiles (Fig. 1a), as would be expected if the plagioclase were more albitic at the surface. We suspect that the decrease in Na counts towards the interior is a result of change-induced migration, as observed by Havette²², and does not reflect real compositional variation in the plagioclase grain. The increase in K near the surface could have a similar explanation²³.

Unweathered feldspar yields nearly constant Al and Si counts from the surface to a depth of ~500 Å. The ratio for ²⁸Si/²⁷Al (Fig. 1a) is constant throughout the profile, with values deviating no more than three per cent from the average. Most importantly, Si/Al values at the freshly cut surface are identical to those for the interior of the same grain. The profile (Fig. 1a) is typical of unweathered grains.

A 1,200 s depth profile of a naturally weathered plagioclase grain is shown in Fig. 1b. Trends for ²⁷Al, ⁴⁰Ca and ³⁹K are similar to those of the fresh plagioclase grain (Fig. 1a). The counts for ²⁸Si, however, are low at the surface compared with

Fig. 2 Percentage change in the $^{28}\text{Si}/^{27}\text{Al}$ ratio of the three weathered plagioclase grains (Fig. 1b) are plotted against depth from the surface of each grain. Si/Al ratios are normalized to Si/Al of the unweathered interior of the same grain, and the percentage change is calculated as described in the text. Symbols: +, fresh plagioclase; $\diamond$, $\times$, *, three samples (1–3 respectively) of weathered plagioclase grains.



the unweathered portion of the same grain (see Fig. 1b, final 200 Å of the profile) or compared with the fresh grain (Fig. 1a).

Normalization minimizes the effects of instrumental variations on count rates, hence ^{28}Si counts have been normalized to ^{27}Al and changes in Si/Al ratios for three weathered plagioclase grains are illustrated in Fig. 2. The ordinate in Fig. 2 is akin to 'permil' of $^{18}\text{O}/^{16}\text{O}$, and differs only in that per cent, rather than 'permil', is plotted. Per cent change in Si/Al = $(R_s/R_N - 1) \times 100$, where R_s is the Si/Al ratio of a sample and R_N is the average value of the Si/Al ratio in the unweathered portion of the same grain. No attempt to standardize these results (wt% basis) is made here. Profiles for the three weathered grains (see Fig. 2, samples 1, 2 and 3) display similar characteristics, but they are very different from the profile for the fresh grain. The ratio for each weathered grain increases continuously with depth until a high constant value is reached (Fig. 2). This high constant value for feldspar sample 3, for example, is 1.256 for the final 200 Å of the profile, whereas the value for the fresh grain is 1.245. Apparently, these high constant values represent fresh unweathered feldspar. We interpret the continuous decrease in the Si/Al ratio towards the surface of the three weathered grains as reflecting a material compositionally different from fresh plagioclase.

Nine feldspar surfaces have been profiled to date. All are similar in that Si/Al increases with depth to a high, constant value for the interior of each grain. Na/Al ratio trends (not shown, see Fig. 1) are similar to Si/Al, but Ca/Al ratio trends (not shown, see Fig. 1) change little as a function of depth (H.W.N. and I.J.M., manuscript in preparation). The trends are consistent and systematic for all nine grains analysed and, although additional studies are required to confirm our hypothesis, the simplest interpretation is that the trends result from compositional variations in an altered surface layer developed on the plagioclase grains.

We now argue that the systematic compositional changes are brought about by soil processes operating at or near the surface of the grains. Evidence for the argument is threefold. Firstly, there is no surface depletion of Si observed for the fresh feldspar. This observation alone suggests that some soil process produced the Si-depleted zones of the grains. Secondly, the grain displaying the lowest Si/Al value at the surface (Fig. 2, sample 3) also has the thickest zone of Si depletion (700–800 Å), and the weathered grain with the highest Si/Al at the surface (Fig. 2, sample 1) displays the thinnest zone of Si depletion (300 Å). Therefore there is a close relationship between Si/Al at the surface and the depth of the Si-depleted layer. As the lowest Si/Al values are at the surface, the process that produced the Si depletion was most effective at the surface of the grain and probably operated at the mineral surface/solution interface. Finally, the history of the weathered plagioclase grains (samples 1–3) and of the fresh feldspar differs in that the surfaces of all the weathered grains were exposed to weathering solutions for

~10,000 years, whereas the surface of the fresh grain has never been exposed to soil solutions. We conclude that the low Si/Al values observed at the surfaces of the weathered grains, and the Si/Al trends (Fig. 2) are the result of preferential removal of Si compared with Al from the near-surface by weathering solutions.

Total dissolved Al is unusually high in some soil waters of the Plastic Lake catchment^{6–10}. The levels often approach or exceed saturation levels for gibbsite and other solid (sometimes amorphous) phases. It has been suggested that these phases might control the Al content of the waters, even though they are scarce or undetected in the soils. Plagioclase and quartz are the dominant minerals of the till from Plastic Lake Catchment. There is a dearth of clay minerals because there is no sedimentary rock component of the till, and few secondary clay minerals have formed *in situ* because the profile is very young (as evidence, cation exchange capacity averages for the soils are less than 1 mequiv. per 100 g). Therefore plagioclase is the ultimate source of dissolved Al in soil waters of Plastic Lake Catchment. Where weathering of plagioclase has occurred, there now exists an Al-enriched layer at its surface; consequently, Al is leached directly from the aluminous surface layer (although derived ultimately from plagioclase). The surface phase is of variable composition, is probably amorphous and thermodynamically unstable relative to aluminous phases such as crystalline gibbsite and boehmite (is more soluble than these minerals). If this aluminous surface phase controls the dissolved Al at or near its saturation value, the solutions must approach saturation, or become supersaturated, with respect to gibbsite and other aluminous minerals. The aluminous surface layer may control Al contents of soil waters in the Plastic Lake and other catchments where plagioclase is an abundant soil constituent.

There is no abrupt or rapid change in Si/Al at the boundary separating the weathered and unweathered zones of the grains (Figs 1b and 2), as expected for boundaries separating compositionally distinct phases. A layer of crystalline precipitates²⁴ was not detected by either SIMS or scanning electron microscope (SEM) studies. These preliminary results provide no evidence that a film of secondary crystalline phases has precipitated on the plagioclase surfaces.

The surfaces of the plagioclase grains have been observed using SEM. Many plagioclase surfaces are free of secondary products and all exposed surfaces display etch pits. Some pits are distributed irregularly over the surfaces of the grains, but others form trains of pits that may mark the loci of dislocations or defects and cleavage traces. The features are similar to those described by Berner and Holdren²⁵ who conclude that a surface reaction is probably the rate-limiting step controlling feldspar dissolution. But secondary residual layers may form and grow, as long as transportation through the altered layer remains rapid relative to the surface reaction. One implication of the SIMS data is that the Al released by feldspar dissolution does not migrate from the site of feldspar dissolution to the same extent

as Si, Ca and Na. Application of congruent and incongruent models to the weathering of plagioclase has to be made with extreme care, and hypotheses more sophisticated than those now available are required to explain these SIMS data and the weathering of plagioclase in nature.

Si/Al trends in the weathered zones (Fig. 2) of the three plagioclase grains result primarily from low Si values (Fig. 1b). The trends represent Si concentration gradients between unreacted plagioclase and solution, and probably chemical potential gradients that promote diffusion of Si towards the solution. The flux of Si may or may not be appreciable, but comment on published diffusion models is warranted. The weathered zone of sample 1, for example, is not substantially leached of Si (composition fairly similar to fresh plagioclase), yet the gradient is nonlinear. Published quasi-steady state models^{26,27} are not viable in that the nonlinear concentration gradients cannot be explained by these models unless diffusion coefficients change systematically throughout the weathered zone. Such changes are likely and must be included in more sophisticated steady-state models.

Recent data¹⁻⁴ indicate that a siliceous residual layer initially forms during dissolution of albite and perthite in acidic solutions. The naturally produced residual layers, in contrast, are Al-enriched. Although the differences may reflect compositional differences in the weathering solutions, these aspects must be fully understood before experimental results can be applied confidently to natural environments.

We thank G. R. Holdren and G. M. Bancroft for critical reading of the manuscript. D. Kirkwood and D. Borre collected the samples, R. L. Barnett provided electron microprobe analyses and SIMS data were obtained at Surface Science Western (UWO). Financial support came from the Ontario Ministry of

the Environment (H.W.N.) and the Natural Sciences and Engineering Research Council of Canada (H.W.N. and G. M. Bancroft, to support I.J.M.).

Received 10 February; accepted 6 June 1988.

1. Chou, L. & Wollast, R. *Geochim. cosmochim. Acta* **48**, 2205-2217 (1984).
2. Chou, L. & Wollast, R. *Am. J. Sci.* **285**, 963-993 (1985).
3. Holdren, G. R. & Speyer, P. M. *Am. J. Sci.* **285**, 994-1026 (1985).
4. Holdren, G. R. & Speyer, P. M. *Geochim. cosmochim. Acta* **51**, 2311-2318 (1987).
5. LaZerte, B. D. & Dillon, P. J. *Can. J. Fish. aquat. Sci.* **41**, 1664-1677 (1984).
6. LaZerte, B. D. *Can. J. Fish. aquat. Sci.* **41**, 766-776 (1984).
7. Hooper, R. P. & Shoemaker, C. A. *Nature* **229**, 463-465 (1985).
8. Sullivan, T. J., Christophersen, N., Muniz, I. P., Seip, H. M. & Sullivan, P. D. *Nature* **323**, 324-327 (1986).
9. Cronan, C. S., Walker, W. J. & Bloom, P. R. *Nature* **324**, 140-143 (1986).
10. Nordstrom, D. K. & Ball, J. W. *Science* **232**, 54-56 (1986).
11. Lovering, J. F. *U.S. natn Bur. Stand., Spec. Publ.* **427**, 135-178 (1975).
12. Shimizu, N., Semet, M. P. & Allegre, C. J. *Geochim. cosmochim. Acta* **42**, 1321-1334 (1978).
13. Liebl, H. *Scanning* **3**, 79-89 (1980).
14. Reed, S. J. B. *Scanning* **3**, 119-127 (1980).
15. Steele, I. M., Hervig, R. L., Hutcheon, I. D. & Smith, J. V. *Am. Miner.* **66**, 526-546 (1981).
16. Shimizu, N. & Hart, S. R. A. *Rev. Earth planet. Sci.* **10**, 483-526 (1982).
17. Hayward, P. J., Hocking, W. H., Doern, F. E. & Cecchetto, E. V. in *Scientific Basis for Nuclear Waste Management* (ed. Lutze, W. L.) 319-328 (North Holland, Amsterdam, 1982).
18. Bancroft, G. M., Metson, J. B., Kresovic, R. A. & Nesbitt, H. W. *Geochim. cosmochim. Acta* **51**, 911-918 (1987).
19. Schott, J. & Petit, J. C. in *Aquatic Surface Chemistry: New Evidence for the Mechanisms of Dissolution of Silicate Minerals* (ed. Stumm, W.) 293-315 (Wiley, New York, 1987).
20. Metson, J. B., Bancroft, G. M., McIntyre, N. S. & Chauvin, W. J. *Surface Interface Anal.* **5**, 181-185 (1983).
21. Metson, J. B., Bancroft, G. M. & Nesbitt, H. W. *Scanning Electron Microscopy* **2**, 595-603 (1985).
22. Havette, A. *Scanning Electron Microscopy* **2**, 585-594 (1985).
23. Streit, L. A., Hervig, R. L. & Williams, P. in *Microbeam Analysis: Quantitative SIMS Microanalysis of Trace Elements in Geological Samples using in situ Ion-implanted Standards* (eds Romig, A. D. & Chambers, W. F.) 91-94 (San Francisco Press, 1986).
24. Helgeson, H. C. *Geochim. cosmochim. Acta* **35**, 421-469 (1971).
25. Berner, R. A. & Holdren, G. R. *Geology* **5**, 369-372 (1977).
26. Luce, R. W., Bartlett, R. W. & Parks, G. A. *Geochim. cosmochim. Acta* **36**, 35-50 (1972).
27. Paces, T. *Geochim. cosmochim. Acta* **37**, 2641-2663 (1973).

Retrospective determination of radon in houses

Christer Samuelsson

Department of Radiation Physics, University of Lund,
University Hospital, S 221 85 Lund, Sweden

In the 1970s it was statistically proved that exposure to radon daughter products caused lung cancer in miners¹. High concentrations of radon daughters have since been found in houses. Any epidemiological radon study begun today is hampered because relevant exposure data are difficult to obtain owing to the long latency period between exposure and tumour manifestation. Here I present a method for measuring cumulative radon daughter levels, which takes advantage of the fact that the first long-lived radon daughter (²¹⁰Pb, half-life 22 yr) becomes firmly attached to glass surfaces in a house. By measuring the surface activity concentration of the alpha-emitter ²¹⁰Po, the time-integrated radon, or radon daughter, concentration can be estimated. Thus, indoor glass can act as a long-term retrospective or prospective exposure meter for radon in dwellings.

Indoor exposure to decay products of radon (²²²Rn) contributes the largest fraction of natural radiation dose to people living in subarctic or temperate regions of the world. Radon is an inert and radioactive gas in the natural decay series that commences with ²³⁸U. As a result of its inert properties, radon is exhaled from the ground and from building materials, for example, and is efficiently trapped in a house. Radon gives birth to decay products (daughters) according to the scheme shown in Fig. 1. Inhaled short-lived radon daughters are caught in the bronchial tree, and it is believed that these alpha-emitting progenies can appreciably increase the risk of lung cancer.

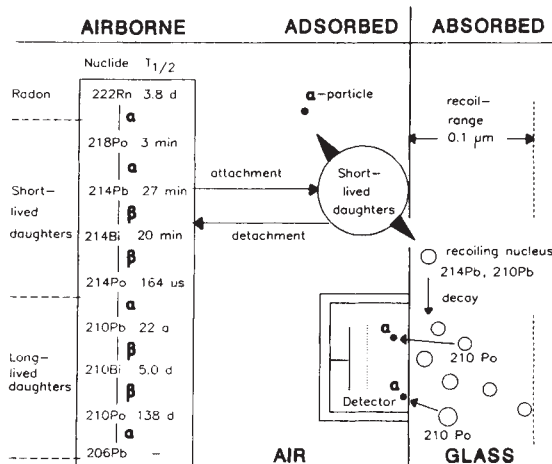


Fig. 1 A schematic illustration of how the radon daughters are attached to a macroscopic surface (glass) by plate-out and how through subsequent alpha decay the progenies are deposited in the superficial layer of the glass. Thus, the glass 'memorizes' the airborne radon activity of several decades before as a result of the long half-life of ²¹⁰Pb. From measurements of the ²¹⁰Po activity of the glass by alpha spectrometry, the accumulated radon concentration can be estimated.

In any epidemiological study an appropriate and exact 'exposure' measure is desirable. In this context, the long delay period between radon exposure and the manifestation of lung cancer in an individual is a serious problem. The idea of using long-lived radon daughters (²¹⁰Pb or ²¹⁰Po) as an estimate of long-term indoor radon concentrations is not new^{2,3}, but so far any such method has failed to provide a quantitative technique for determining the amount of radon in houses. ²¹⁰Po and ²¹⁰Pb

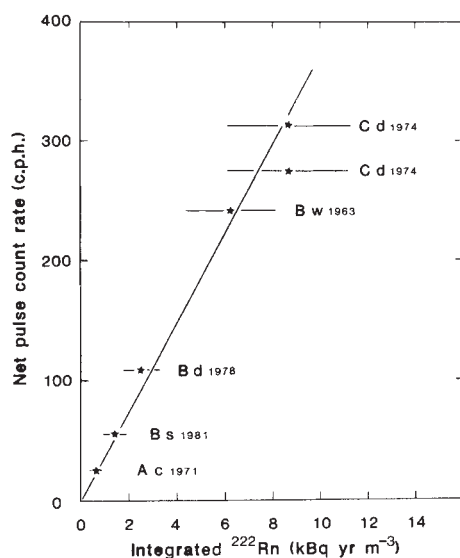


Fig. 2 The net pulse count rate (in counts per hour) of the ^{210}Po alpha peak as a function of integrated ^{222}Rn concentration (kBq yr m^{-3}) for glass samples taken from three different dwellings (A, B and C). c denotes glass taken from a framed photograph, d is taken from an interior door, w is a sample from a window facing the interior of the house and s is a marble ornament. The year of the start of radon exposure is given after each point. Glass samples with an area of 210 cm^2 were analysed in a pulse ionization chamber. (A count rate of 100 c.p.h. corresponds roughly to a surface activity of 3 Bq m^{-2} .)

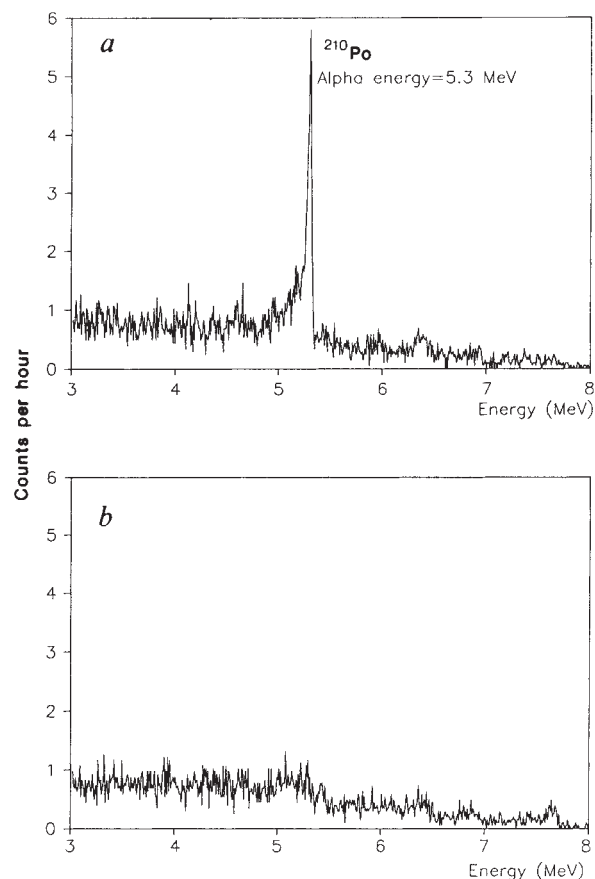


Fig. 3 Alpha spectra from the glass sample A c 1971 (Fig. 2) obtained in a pulse ionization chamber. a, Side facing the room and b, side facing the photograph. The radionuclide ^{210}Po is detectable only on the side facing the room. The narrowness of the peak indicates that the activity must be located in the outermost layer of glass only.

have been used as surrogates for long-term ^{222}Rn in the outdoor environment⁴. Here I show that a combination of vitreous glass as substrate, alpha-emitting ^{210}Po as tracer, and a pulse ionization chamber as detector can provide a retrospective radon exposure system (see Fig. 1). As a result of the long half-life (22 yr) of ^{210}Pb , the grandparent of ^{210}Po , we can obtain information on integrated radon levels in a house over the past decades.

These findings are also important for the future. The glass-polonium system can act as a prospective personal radon exposure meter in the form of, for example, a small mirror used during a person's lifetime. By analysing the mirror surface, say every other year, the accumulating radon exposure can be followed. Spectacles provide a portable personal meter, but the correlation between surface ^{210}Po activity and accumulated radon concentration may be poor because of variation in such habits of cleaning and storing. Also the small area means that the ability of the pulse ionization chamber to analyse large surfaces cannot be fully exploited. The potential ability of spectacles as retrospective radon doseimeters has also been pointed out by Fleischer⁵. His suggestion is based on the track-etch technique and is limited to spectacles with plastic lenses that give etchable alpha tracks.

The kinetic energy transferred to ^{210}Pb from alpha decay of ^{214}Po to ^{210}Pb causes the ^{210}Pb atom to recoil into the superficial layer of the glass. As the recoil range of the atom in glass is at most about $0.1 \mu\text{m}$ (0.0001 mm), an almost ideally thin alpha source is formed on the subsequent alpha decay of ^{210}Po (Fig. 1). Alpha spectrometric analysis of glass surfaces of different ages indicates that diffusion processes are insignificant in displacing long-lived radon progeny atoms in glass. The plate-out on the glass of short-lived radon daughters is the chief source of ^{210}Pb and ^{210}Po in the upper layer of the glass. As the plate-out process is fairly rapid (rate constants $\geq 1 \text{ per h}$) the ^{210}Po surface activity of the glass is fairly insensitive to cleaning, providing the frequency of cleaning is less than once a day.

The results from glass samples analysed so far are plotted in Fig. 2 as a function of exposure from radon. The house identification (A to C), the type of glass (c, s, d, or w) and the year in which exposure began are given for each point. Exposures were interrupted during February or March 1988 and the glass samples were analysed for about 24 h in a pulse ionization chamber. (This chamber was built during the author's stay at the Alligator Rivers Region Research Institute, Jabiru, Australia, in October 1987, and is of a type similar to the one developed by Hötzel and Winkler in West Germany⁶.) Glass samples were cut as a square of area $\sim 210 \text{ cm}^2$. No decay correction has been applied to the pulse count values in Fig. 2. The alpha spectra from the least exposed sheet of glass (A c 1971) in Fig. 2 are shown in Fig. 3. The side protected from room air shows no traces of ^{210}Po , thus proving that the original source must be located outside the glass itself.

The exposure estimates in Fig. 2 are all based on at least one long-term (weeks or months) measurement of ^{222}Rn or radon daughters, and all detector systems used are traceable to the international standards through intercomparison exercises in the UK (ref. 7). But the exposure estimates are still uncertain because of the non-continuous type of measurements, and also because the radon or radon daughter detectors were situated in most cases in rooms other than those from which the glass samples were taken. To accommodate this uncertainty, each exposure value has been assigned a somewhat arbitrary error of $\pm 30\%$. Another source of inaccuracy inherent in this method is that it is a one-parameter measurement describing conditions governed and influenced by many parameters, such as individual

short-lived daughter levels and turbulence in the room air and aerosol size distribution. Despite our oversimplification, the results in Fig. 2 show that the correlation between radon in air and polonium in glass is good.

From Fig. 2 it is seen that the two sides of the door glass from house C have different polonium activities, indicating that the radon levels were different in the two rooms separated by the door. Unfortunately, this hypothesis could not be tested as countermeasures against radon have recently been taken in house C, but a room-to-room variation in house C is plausible because the radon in the house stems from the infiltration of ground air along water supply pipes and electrical cabling.

The polonium activity in the glass samples most exposed (Fig. 2) can be detected by the widely available surface barrier detector, but its sensitivity is significantly less than that of the pulse ionization chamber because of the much smaller detector area. Simple total-alpha detectors (such as ZnS detectors) may also be sensitive enough for the most exposed glass samples, but can be subject to interference from alpha-emitting nuclides other than ^{210}Po .

Received 2 May; accepted 22 June 1988.

1. Lundin, F. E., Wagoner, J. K. & Archer, V. E. *Joint Monograph 1* (US Department of Health, Education and Welfare, National Institute for Occupational Safety, and National Institute of Environmental Health Sciences, 1971).
2. Lively, R. S. & Ney, E. P. *Health Phys.* **52**, 411-415 (1987).
3. Forberg, S. *Final report project SSI P 327.86* (Royal Inst. of Technology, Stockholm, 1987).
4. Turekian, K. K., Nozaki, Y. & Benninger, L. K. *A. Rev. Earth planet. Sci.* **5**, 227-255 (1977).
5. Fleischer, R. L. *Health Phys.* **52**, 219-221 (1987).
6. Hötzel, H. & Winkler, R. *Nucl. Instrum. Meth.* **150**, 177-181 (1978).
7. Miles, J. C., Stares, E. J., Cliff, K. D. & Sinnaeve, J. *Radiat. Prot. Dosimetry* **7**, 169-173 (1984).

A novel free-living prochlorophyte abundant in the oceanic euphotic zone

Sallie W. Chisholm, Robert J. Olson*, Erik R. Zettler*, Ralf Goericke†, John B. Waterbury* & Nicholas A. Welschmeyer†

48-425 Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

* Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

† Harvard University, Cambridge, Massachusetts 02138, USA

The recent discovery of photosynthetic picoplankton has changed our understanding of marine food webs¹. Both prokaryotic^{2,3} and eukaryotic^{4,5} species occur in most of the world's oceans and account for a significant proportion of global productivity⁶. Using shipboard flow cytometry, we have identified a new group of picoplankters which are extremely abundant, and barely visible using traditional microscopic techniques. These cells are smaller than the coccoid cyanobacteria and reach concentrations greater than 10^5 cells ml^{-1} in the deep euphotic zone. They fluoresce red and contain a divinyl chlorophyll *a*-like pigment, as well as chlorophyll *b*, α -carotene, and zeaxanthin. This unusual combination of pigments, and a distinctive prokaryotic ultrastructure, suggests that these picoplankters are free-living relatives of *Prochloron*⁷. They differ from previously reported prochlorophytes—the putative ancestors of the chloroplasts of higher plants—in that they contain α -carotene rather than β -carotene and contain a divinyl chlorophyll *a*-like pigment as the dominant chlorophyll.

In recent cruises in the North Atlantic and Pacific we discovered a flow cytometric 'signature' from extremely abundant red-fluorescing cells deep in the euphotic zone, below the *Synechococcus* maximum (Fig. 1a). These picoplankters typically have forward light scatter signals (which are related to cell size) less than a third of those of *Synechococcus* cells in the

same sample. Most pass through a 0.8- μm Nuclepore filter and are retained by a 0.6- μm filter. They do not fluoresce in the orange region of the spectrum (540 to 630 nm), indicating the absence of phycoerythrin^{8,9}. As in *Synechococcus*, both forward light scatter and fluorescence per cell increase with depth in these cells, presumably reflecting acclimation to decreasing light levels (Fig. 1b). The cells are very difficult to detect by epifluorescence microscopy; only in the deepest samples do microscopic counts approach those obtained by flow cytometry.

We have mapped and characterized these unusual cells on several cruises since 1985 in the southern California Bight, Panama Basin, Gulf of Mexico, Caribbean, and North Atlantic between Woods Hole, Massachusetts and Dakar, Senegal. Of the 67 stations we sampled, the cells were undetectable only in shallow, well-mixed waters on Georges Bank off Cape Cod, Massachusetts, and in the northern Sargasso Sea in May, when stratification had only recently become established. In all other instances their flow cytometric signature was clearly defined in the deep euphotic zone, although it usually disappeared into the baseline noise in samples collected from above the 5% isolume. Maximum cell concentrations at the majority of the stations ranged from 5×10^4 to 1.2×10^5 cells ml^{-1} . Only 10 of the stations had fewer, and these tended to be coastal locations. The depth at which the cells were easily detectable and maximally abundant was closely related to the depth of the primary nitrite maximum layer, a chemical marker for the bottom of the euphotic zone¹⁰. The *Synechococcus* population was always located above the nitrite maximum layer (see Fig. 1b), which was always 20-100 m below the top of the thermocline.

Wet mounts from a cell concentrate collected from 100 m and examined by light microscopy, revealed small coccoid to rod-shaped cells under phase contrast illumination that fluoresced red when examined by epifluorescence microscopy using the Zeiss filter set 48 77 05. Transmission electron micrographs of samples from the same cell concentrate revealed an abundance of prokaryotic cells that were smaller and more numerous than their *Synechococcus* neighbours, and had closely appressed peripheral thylakoids (Fig. 2b). These cells are identical in appearance to those described by Johnson and Sieburth² as type II *Synechococcus* which they also found to be numerically dominant at 100 m in the Sargasso Sea. These authors noted that the fluorescence properties of the cells suggested that they might not contain phycoerythrin. The tight packing of the thylakoids in these cells is similar to that seen in prochlorophytes¹¹, and rather distinct from *Synechococcus*. The latter have widely spaced thylakoids (40-50 nm) accommodating phycobilisomes on the outside of the membranes¹² (Fig. 2a). The ultrastructure of the cell shown in Fig. 2b, especially the arrangement of the cytoplasmic membrane system, also resembles that of nitrifying bacteria belonging to the genus *Nitrosomonas*¹³. But marine representatives of *Nitrosomonas* possess a sculptured outer cell wall layer¹³ which is lacking in our cells, and their reported concentrations¹⁴ for the open ocean are of the order of 50 cells per ml.

Dual beam flow cytometric analysis⁹ of the cells, in addition to indicating the absence of phycoerythrin, revealed that the relative intensity of their red fluorescence (660-700 nm) when excited by blue light (488 nm) was about six times that when excited by green light (515 nm). This suggests the absence of phycocyanin and photosynthetically active carotenoids. Analysis of the cells' pigment composition by HPLC gave a unique combination of pigments (Tables 1 and 2): A new type of chlorophyll *a*, probably divinyl chlorophyll *a* (see footnote to Table 1), chlorophyll *b*, zeaxanthin, α -carotene and traces of another uncharacterized pigment, possibly chlorophyll *c* or a chlorophyllide. Lutein, β -carotene, and prasinoxanthin were not detected and normal chlorophyll *a* was present in traces, constituting only 1.7% of the total chlorophyll-*a* fraction. The absence of xanthophylls other than zeaxanthin, especially prasinoxanthin or lutein, is consistent with our observation that

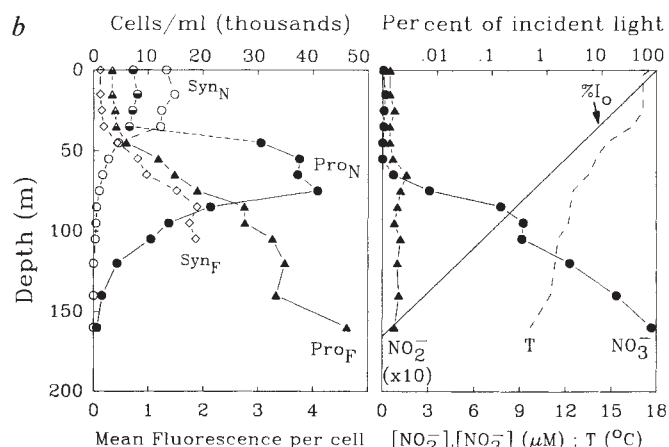
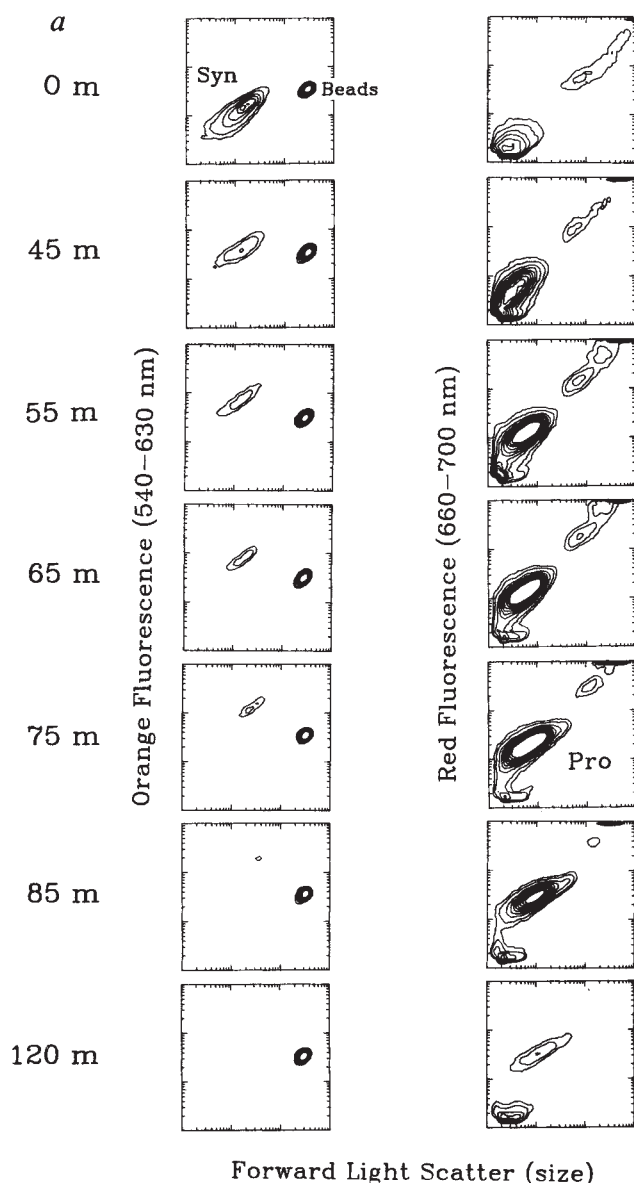


Fig. 1 *a*, Flow cytometric signatures of *Synechococcus* and the novel prochlorophytes from a depth profile off Southern California (33°11.73' N, 118°29.82' W; 14 December 1987). Data from measurements of up to 50,000 individual cells per sample are presented as two-dimensional flow contour plots. Measurements were made using a Coulter Epics V flow cytometer⁹. Left, forward light scatter versus orange fluorescence (from phycoerythrin) reveals *Synechococcus* populations (Syn) in the upper waters. Fluorescent beads (0.9-μm diameter, Duke Scientific) were added as internal standards. Right, forward light scatter versus red fluorescence reveals very small cells (Pro) dominating the deeper samples. The less dense cloud of larger cells with brighter fluorescence is a mixture of eukaryotic phytoplankton cells about 2–5 μm in diameter. Signals from *Synechococcus* cells and the standard beads are not visible here because the data have been gated to show only signals from particles having no orange fluorescence. *b*, Depth profile of biological and hydrographic features at the station described in *a*. Temperature (*T*) was obtained from an expendable bathythermograph; light penetration (% *I*₀) was calculated from Secchi depth according to (*I*_z = *I*₀*e*^{−*kz*}, where *I*_z = the light intensity at depth *z*, *I*₀ = incident light intensity, and *k* = 1.7/Secchi depth); nitrate and nitrite (NO₃[−], NO₂[−]) were analysed by standard techniques²⁴. Samples in which we were able to see only part of a population in the flow cytometric signature are indicated by half-filled symbols and dashed lines. Mean fluorescence per cell (Syn_F, Pro_F) from flow cytometric signatures are expressed relative to uniform standard beads for each emission band⁸. Cell concentrations (Syn_N, Pro_N) were calculated as described in Olson *et al.*⁹.

Table 1 Concentrations of pigments in the prochlorophytes and their absorption maxima (*A*_{max}) compared to those reported in the literature. The samples were from 100-m depth in the Gulf Stream (38°09.5' N, 66°50.8' W; 19 September 1987)

Pigment	Concentration* (fg cell ^{−1})	<i>A</i> _{max} † this study (nm)	<i>A</i> _{max} reported (nm)	Reference
Chlorophyll <i>a</i>	trace‡	—	428, 660	22, 23
Pheophytin <i>a</i>	—	408, 667‡	408, 667	22, 23
Divinyl chlorophyll <i>a</i>	2.15‡	—	436, 661	22, 23
Divinyl pheophytin <i>a</i>	—	417, 667‡	417, 667	22, 23
Chlorophyll <i>b</i>	2.31	457, 646	455, 645	17
Zeaxanthin	0.69§	(429), 450, 478	(425), 450, 478	17
α-Carotene	0.29	(423), 444, 475	423, 444, 473	28

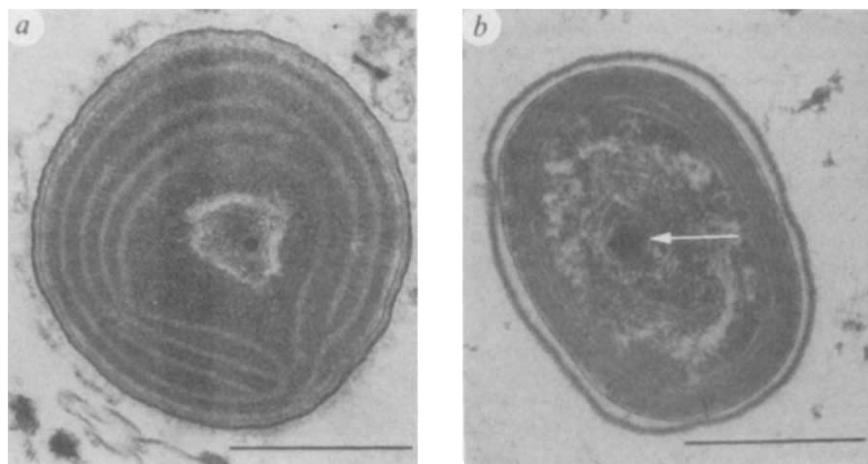
* Cellular concentrations of pigments are means from five filter-fractionated (0.6–0.8 μm) samples extracted in 100% acetone and analysed by reverse phase HPLC on a Microsorb C-18, 3-μm column using a slightly modified version of the method of Mantoura and Llewellyn²⁷. The samples were gravity-filtered through Nuclepore filters, and the resulting fraction was analysed by flow cytometry as well. The fractions consisted of 98% prochlorophytes according to their integrated light scatter and fluorescence signals, thus they were considered a 'pure' sample for our purposes.

† Absorption spectra were obtained on pigments from a much larger sample obtained by tangential flow filtration of 100 l of whole seawater from 70 m at the same station. The chlorophyll *a* pigments and their derivatives were analysed in diethyl ether; chlorophyll *b* and zeaxanthin in acetone, and α-carotene in ethanol. Numbers in parentheses indicate the presence of a shoulder rather than a peak in the spectrum.

‡ Chlorophyll *a* and the divinyl chlorophyll-*a*-like pigment are not fully resolved on the reverse-phase HPLC system used, however, their respective pheophytins are. Thus we collected the two coeluting pigments, converted them to their respective pheophytins by acidification with 1N HCl and isolated the two pheophytins. The absorption maxima of the new type of pheophytin-*a* are identical to those published by Bazzaz^{22,23} for divinyl pheophytin-*a*, thus we hypothesize that the new chlorophyll *a* in our samples is divinyl chlorophyll *a*.

§ Because zeaxanthin and lutein coelute on this system, we confirmed the absence of lutein in the samples by analysing the zeaxanthin fraction on a reverse-phase isocratic tetrahydrofuran:H₂O (45:55% by volume) system which separates the two pigments.

Fig. 2 Electron micrographs of thin sections of marine *Synechococcus* (a) and prochlorophyte (b). Cells were concentrated by tangential flow filtration from 600 l of water collected at 100 m at 36°07.9' N, 64°18.1' W, 11 September 1987. The primary distinction between the two organisms is the arrangement of the photosynthetic thylakoids. Note the carboxysomes (arrow) in the prochlorophyte indicating that it contains one of the key autotrophic enzymes (RUBP carboxylase). Flow cytometric analysis of the cell concentrate before fixation indicated that prochlorophytes outnumbered *Synechococcus* by a factor of 100; in the thin sections examined, the ratio of prochlorophytes to *Synechococcus* was 5 ($n = 50$). We attribute the differences to losses of the prochlorophytes during fixation. Scale bar, 0.5 μm .



the cells are prokaryotic^{15,16}, and the presence of chlorophyll *b* confirms that they belong in the *Prochlorophyta*¹¹. The presence of α -carotene is puzzling, because it is generally assumed that prokaryotes are unable to synthesize carotenoids with epsilon-rings^{15,17}, like α -carotene. We expect that some β -carotene was present as a metabolic intermediate, but was below our limit of detection. The chlorophyll *a/b* ratio measured for these cells is approximately 1, which contrasts sharply with that of *Prochloron* (4–7)¹⁸ and *Prochlorothrix* (8–9)¹⁹ (Table 2), but is similar to values reported for marine chlorophytes (1–3)²⁰.

Perhaps the most unusual pigment characteristic of the cells is the divinyl chlorophyll-*a*-like molecule. This pigment elutes earlier than chlorophyll *a* and is characterized by a red-shift of the Soret peak by 8–10 nm compared to normal chlorophyll *a*. Gieskes and Kraay²¹ isolated a pigment very similar to this from the less-than-one μm fraction in surface waters of the tropical Atlantic. The absorption maxima for this pigment (436 and 661 nm in diethyl ether) are identical to those reported by Bazzaz^{22,23} for divinyl chlorophyll *a*.

At several stations we were able to obtain estimates of the primary production attributable to the prochlorophytes, using the carbon-14 method^{24,25} coupled with post-incubation size fractionation (collecting the 0.4–0.8- μm fraction using Nuclepore filters). These data were corroborated by sorting the cells in question after incubation with ¹⁴C-labelled bicarbonate, using the cell-sorting capability of the flow cytometer²⁶. The proportion of total productivity below the 1% isolume attributable to these cells ranged from 15 to 60%; absolute rates of production ranged from 7×10^{-3} to $1.2 \times 10^{-1} \mu\text{g C l}^{-1} \text{h}^{-1}$. If we assume that the cells are 0.8- μm -diameter spheres with the same relative C content as cultured *Synechococcus*¹² (200 fg μm^{-3}), the results of our carbon fixation experiments translate into doubling times ranging from 2 to 10 days.

These prochlorophytes seem well adapted for life at the bottom of the euphotic zone; high concentrations of chlorophyll *b* as well as the red shift of the Soret peak of the divinyl chlorophyll-*a*-like pigment optimizes absorption of which blue light (460–480 nm) dominates the deep ocean. We have been unable to adequately quantify the prochlorophytes in the upper waters of the euphotic zone with the flow cytometer because of their reduced fluorescence and light scatter at these depths (Fig. 1a). But the divinyl chlorophyll-*a*-like molecule associated with these cells has been detected in the surface layers of the tropical and subtropical Atlantic²¹, suggesting that they are sometimes abundant in the upper waters of the oceans.

Unfortunately, our attempts to isolate these cells into pure culture have failed. Johnson and Sieburth² reported similar difficulties with their Type II prokaryotes, although they had no difficulties isolating *Synechococcus* from the same waters. Although unequivocal identification of our red-fluorescing picoplankters awaits their isolation and nucleotide sequencing, the evidence presented here indicates that they represent a new type of chlorophyll-*b*-containing prokaryote, that is, a Prochlorophyte (*sensu* Lewin¹¹). They differ from other members of this group in that they contain a different type of chlorophyll *a*, α -carotene rather than β -carotene, and the chlorophyll *a/b* ratio of cells found at the bottom of the euphotic zone is significantly lower than that of *Prochloron* or *Prochlorothrix* (Table 2). If our identification is correct, they represent the first free-living marine prochlorophyte, and elevate the status of this division to one of considerable importance in marine ecosystems. Regardless of taxonomic affinity, their ubiquity and abundance indicates that they are a significant component of the microbial food web and potentially important primary producers in temperate and tropical oceans.

We thank E. V. Armbrust, M. Gerath, H. Sosik, A. Lewitus and F. Valois for their assistance and the captains and crews of

Table 2 Comparison of the pigment composition of the prochlorophyte with its prokaryotic relatives^{16,17,19}

Pigment	<i>Prochloron</i>	<i>Prochlorothrix</i>	<i>Synechococcus</i>	New prochlorophyte
Chlorophyll <i>a</i>	+	+	+	—
Divinyl chlorophyll <i>a</i> -like pigment	—	—	—	+
Chlorophyll <i>b</i>	+	+	—	+
Chlorophyll <i>a/b</i> ratio	4–7	8–9	—	1*
Phycobiliproteins	—	—	+	—
α -Carotene	—	—	—	+
β -carotene	+	+	+	—
Zeaxanthin	+	+	+	+

* Note that this ratio was measured on cells found deep in the euphotic zone, and could change in cells grown in high light.

the RVs Gyre, Oceanus, Knorr and Atlantis II for their cooperation. This work was supported by grants from the US National Science Foundation and Office of Naval Research.

Received 18 May; accepted 10 June 1988.

1. *Photosynthetic Picoplankton* (eds Platt, T. & Li, W. K. W.) *Can. Bull. Fish. Aquat. Sci.* Vol. 214 (1986).
2. Johnson, P. W. & Sieburth, J. McN. *Limnol. Oceanogr.* **24**, 928-935 (1979).
3. Waterbury, J. B., Watson, S. W., Guillard, R. R. L. & Brand, L. E. *Nature* **277**, 293-294 (1979).
4. Johnson, P. W. & Sieburth, J. McN. *J. Phycol.* **18**, 318-327 (1982).
5. Murphy, L. S. & Haugen, E. M. *Limnol. Oceanogr.* **30**, 47-58 (1985).
6. Platt, T., Subba Rao, D. V. & Irwin, B. *Nature* **301**, 702-704 (1983).
7. Lewin, R. A. & Withers, N. W. *Nature* **256**, 735-737 (1975).
8. Olson, R. J., Vulot, D. & Chisholm, S. W. *Deep Sea Res.* **32**, 1273-1280 (1985).
9. Olson, R. J., Chisholm, S. W., Zettler, E. R. & Armbrust, E. V. *Deep Sea Res.* **35**, 425-440 (1988).
10. Herbland, A. & Voituriez, B. *J. mar. Res.* **37**, 87-101 (1979).
11. Lewin, R. A. in *The Prokaryotes* Vol. 1 (eds Starr, M. P., Stolp, H., Truper, H. G., Balows, A. & Schlegel, H. G.) 257-266 (Springer, Berlin, 1981).
12. Waterbury, J. B., Watson, S. W., Valois, F. W. & Franks, D. G. in *Photosynthetic Phytoplankton* (eds Platt, T. & Li, W. K. W.) 71-120 (*Can. Bull. fish. aquat. Sci.*, Ottawa, 1986).
13. Watson, S. W., Valois, F. W. & Waterbury, J. B. in *The Prokaryotes* (eds Starr, M. P., Stolp, H., Truper, H. G., Balows, A. & Schlegel, H. G.) 1005-1022 (Springer, Berlin, 1981).
14. Ward, B. B. *J. mar. Res.* **40**, 1155-1172 (1982).
15. Goodwin, T. W., *The Biochemistry of Carotenoids* Vol. 1, second edition (Chapman and Hall, London, 1980).
16. Guillard, R. R. L., Murphy, L. S., Foss, P. & Liaaen-Jensen, S. *Limnol. Oceanogr.* **30**, 412-414 (1985).
17. Foss, P. R. A., Lewin, S. & Liaaen-Jensen, S. *Phycologia* **26**, 142-144 (1987).
18. Withers, N. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 2301-2305 (1978).
19. Burger-Wiersma, T., Veenhuis, M., Korthals, H. J., Van de Wiel, C. C. M. & Mur, L. R. *Nature* **320**, 262-264 (1986).
20. Wood, A. M. *J. Phycol.* **15**, 330-332 (1979).
21. Gieskes, W. W. & Kraay, G. W. *Limnol. Oceanogr.* **28**, 757-766 (1983).
22. Bazzaz, M. B. *Photobiochem. Photobiophys.* **2**, 199-207 (1981).
23. Bazzaz, M. B. & Brereton, R. G. *FEBS Lett.* **138**, 104-108 (1982).
24. Strickland, J. D. H. & Parsons, T. R. *A Practical Handbook of Seawater Analysis*, second edition (Bull. 167, Fish. Res. Bd. Can., Ottawa, 1972).
25. Fitzwater, S. E., Knauer, G. A. & Martin, J. H. *Limnol. Oceanogr.* **27**, 544-551 (1982).
26. Li, W. K. W. in *Photosynthetic Phytoplankton* (eds Platt, T. & Li, W. K. W.) 251-286 (*Can. Bull. fish. aquat. Sci.*, Ottawa, 1986).
27. Mantoura, R. F. C. & Llewellyn, C. A. *Analyt. chim. Acta* **151**, 297-314 (1983).
28. Davies, B. H. in *Chemistry and Biochemistry of Plant Pigments* Vol. 2 (ed. Goodwin, T. W.) 38-165 (Academic, London, 1976).

Mineral nutrition and spatial concentrations of African ungulates

S. J. McNaughton

Biological Research Laboratories, Syracuse University, Syracuse, New York 13244-1220, USA

Africa's abundant large herbivores are very heterogeneously distributed, both geographically and regionally¹. Within a region, some localities contain dense animal concentrations although areas nearby may be virtually unoccupied. Mixed-species herds are a conspicuous feature of areas where animals concentrate²⁻⁴. The prevailing explanations of local distributional concentrations are (1) that different herbivore species facilitate each other's foraging, and (2) that animals are protected from predation by both intra-specific and interspecific association⁵⁻⁹. If facilitation of grazing were an overriding factor, mixed species herds should move extensively with localized rain showers to obtain the greatest forage yield¹⁰. If predation were the major factor influencing animal densities and distributions, rapid, unpredictable spatial movements would further reduce predation. But because resident, non-migratory species tend to occupy home ranges that are stable over time¹¹, neither of these hypotheses is totally compelling. Because tropical forages are of lower quality than temperate ones and are often chronically deficient in mineral elements¹², I tested the hypothesis that areas where animals concentrate are localities supporting forages of higher mineral content. I report here that the mineral content of foods is an important determinant of the spatial distributions of animals within the Serengeti National Park, Tanzania. Based on ecological criteria, magnesium, sodium and phosphorus appear particularly important.

Areas supporting high densities of animals were identified using data and observations collected since 1974 in the course of studies of the grazing ecosystem in the park¹³⁻¹⁵ and from the records of the Serengeti Ecological Monitoring Programme, mainly concentrated from 1969 to 1978 (ref. 16). No locations were sampled on the Serengeti Plains because they are not occupied by herbivores all year around. During the 1986 wet season, samples of the youngest leaf blades on actively growing grasses were collected from grasslands on and adjacent to areas of high and low animal density, dried, returned to Syracuse University, and analysed by inductively coupled plasma spectrometry using standard sample preparation and analytical protocols¹⁷⁻¹⁹. Soil samples from the upper 10 cm immediately below sampled grasses were collected simultaneously and treated with similar protocols. Data for the 19 elements determined in forage samples were analysed for location by discriminant analysis followed by a Kruskal-Wallis rank analysis to determine which elements differed significantly between the different localities. Soil samples were compared by the Kruskal-Wallis test for elements discriminating between forage samples.

Discriminant analysis indicated that the mineral contents of forages differentiated areas of high and low animal density with a high degree of accuracy. Only two of 33 control and two of 36 from areas of high animal density were misclassified ($\chi^2 = 69.9$, $P < 0.000001$; canonical correlation = 0.843). Ten of the 19 elements differed significantly between control forages and those supporting a high animal density (Table 1). The largest proportional difference was in Na concentrations, which averaged over three times higher in the latter forages. Concentrations of Al and Fe were over 80% higher in these forages; P concentrations were 50% greater, Mn and Pb 40% greater, and Ca, Mg and V 10-23% greater. Ni was much lower in forages supporting high animal densities. Elements important to plants but of minor, if any, importance to animals, such as B, did not differ between localities.

Underlying edaphic and geological properties can contribute to differences in plant mineral concentrations²⁰⁻²¹. But none of the soil variables aided in interpreting plant variation (Table 2). Thus, ecological factors other than the underlying differences in general physical environments govern the mineralogical differences between forages from grasslands supporting high and low animal densities.

Nutrient availability to both plants from environmental pools and to herbivores from plant tissues is determined by many complex factors involving both the biotic and the abiotic components of ecosystems²²⁻²⁶. Mineral concentrations in plants are a complex function of environmental variables as well as such plant properties as species, ecotype, growth stage, tissue and growth rate²⁷, many of which will vary between the grazed grasslands of areas of high animal density and the ungrazed areas sampled here as controls^{10,13-15}. But the restriction of samples to the youngest leaves of actively growing grasses would have stabilized many potential variables²²⁻²⁷. Animal nutritional status will also depend upon a wide variety of factors, including the plant variables above, as they affect intake and digestibility, as well as other complex environmental and animal properties²⁴⁻²⁶. But forage mineral analysis is a reliable index of the general ability of forages to meet animal mineral needs¹², and the data reveal several potentially important nutritional differences between forages on areas of high animal density and control locations. Using beef cattle feeding standards²⁸ as an approximate index of nutritional requirements of ruminants, of those elements distinguishing forages from the two types of localities, on control locations Mg and Na were below general standards, whereas P was below the requirements for lactating cows and growing animals. There can be little doubt that Na is an important component of the nutrition of wild animals²⁹⁻³² and the data indicate that grazers on the Serengeti can meet their dietary Na needs solely from forages with high nutrient concentration, an unusual occurrence in animal nutrition^{12,24-26}. Based on

Table 1 Discriminant analysis and significance testing for elemental differences between forage samples ($n = 69$) from grasslands lacking animal concentrations (controls) and with animal concentrations

Element	Standardized discriminant function coefficient	Concentrations (p.p.m.)		K-W	P
		Controls	Areas of high animal density		
Al	0.754	137	257	26.19	0.000001
B	-0.019	5.93	6.23	0.79	NS
Ca	0.259	3,447	4,056	7.83	0.005
Cd	-0.194	0.025	0.031	0.86	NS
Co	-0.098	0.238	0.266	0.10	NS
Cr	0.688	0.80	1.49	0.02	NS
Cu	0.298	5.37	6.29	1.17	NS
Fe	0.178	96.6	178.2	25.21	0.000001
K	0.560	19,880	20,533	<0.01	NS
Mg	-0.273	1,602	1,960	5.62	0.032
Mn	0.491	55.5	77.0	3.65	0.056
Mo	-0.051	2.05	2.02	0.06	NS
Na	0.629	894	3,097	5.60	0.018
Ni	-0.496	0.99	0.18	4.42	0.036
P	0.873	2,772	4,162	6.98	0.0082
Pb	0.127	3.14	4.14	4.02	0.044
Se	-0.039	0.809	0.915	0.93	NS
V	0.061	1.58	1.83	7.31	0.007
Zn	-0.347	18.2	16.3	1.07	NS

Centroid values are -1.61 for controls and 1.48 for areas of high animal density. Concentrations are parts per million (p.p.m.) on a dry weight basis. K-W is the Kruskal-Wallis statistic for elemental differences and P is the associated probability level (NS = not significant). Samples were collected from the most abundant grass species in a sward except when there were coequal dominants and both species were collected; thus sample size between areas with high animal density and controls were not exactly equivalent. Control samples were within 5 km of the centres of areas of high animal density in all cases. Samples were collected into plastic bags, sun-dried in plastic mesh bags, and after return to Syracuse were washed in double distilled water, redried, ground in a Wiley mill with a stainless steel grinding chamber, and dry ashed at 500 °C for 12 h. All labware was washed with acid and solutions were prepared with water that was distilled and subsequently passed through a four-column Barnstead Nanopure II system. Analysis of the water indicated undetectable or negligible quantities of elements assayed. Ashed samples were taken up in 2 M HCl spiked with Y at 100 p.p.m. for an internal standard. Analyses were done in simultaneous mode on a Leeman Labs PlasmaSpec 2.5 Inductively Coupled Argon Plasma Spectrometer. All statistical analyses are done with the STSC Statgraphics statistical analysis system running under Disk Operating System version 3.10 on an IBM PC-AT computer.

standardized discriminant function coefficients, their importance in animal nutrition, and the relative magnitude of differences between control and experimental forage samples, Mg, Na and P appear to be particularly important factors in Serengeti grassland ecology.

Mineral contents could be correlated with some other forage variable that ungulates seek in their diet. To test this, an independent forage set ($n = 56$) previously subjected to standard forage analysis³³ was analysed for mineral content. Mineral content was then subjected to multivariate statistical analysis with forage N, a factor frequently emphasized in wildlife nutrition²⁵. Only Cu ($P = 0.0001$), K ($P = 0.001$) and Zn ($P = 0.005$) were significantly correlated with N, and they were not among the elements discriminating areas of high animal density from control locations.

Tropical forages are so notoriously poor in mineral elements that complete supplementation is advocated as a standard husbandry practice for domestic livestock in the tropics¹². Veterinary studies have shown that wild ungulates in Africa can also suffer from the same nutritional diseases that plague unsupplemented livestock there³⁴⁻³⁷. Nevertheless, it is obvious that the Serengeti ecosystem, which supports over 2.5 million head of large herbivores, is nutritionally sufficient to meet those animals' needs. Evolution of the animals, with nutritional requirements that may differ from domestic animals, could play a role. Mineral licks²⁹⁻³² also can satisfy certain nutritional deficits in the diet, but many areas supporting high animal densities contain no evident licks and the forage data indicate that such supplementation would not be required for the elements assayed here.

The data suggest two additional hypotheses about the spatial distributions of African animals that would be amenable to testing with appropriate data sets. First, leks, small and highly aggregated mating territories of males of some African antelopes¹¹, could be a consequence of the underlying mineral

properties of forages, for example, high Na concentrations, that tend to attract females. Second, the tendency for male antelopes to stand upon termite mounds¹¹ might also be due to mineral enrichment of forages there by the subsoil excavated to the surface by termites, and a similar attractiveness to females of mineral-enriched forages.

I do not intend to suggest that protection from predators and grazing facilitation are not also involved in ungulate spatial concentrations⁵⁻⁹. In particular, the tendency for prey to concentrate in buffer zones between territorial predators³⁸⁻³⁹ has never been studied in African ecosystems, to my knowledge. Nevertheless, the distinctive mineralogical properties of forages from areas supporting high animal densities suggest that localized areas of forage mineral sufficiency, a hitherto overlooked ecological factor, may be an important component of both animal

Table 2 Mean soil concentrations of elements discriminating controls from areas of high animal density based on forage analysis

Element	Controls	Areas of high animal density
Al	4,760	4,821
Ca	2,577	2,024
Fe	3,330	4,226
Mg	529	377
Mn	250	313
Na	180	162
Ni	8.5	8.4
P	449	481
Pb	42	50
V	6.6	7.2

The concentrations of elements are in p.p.m. None of the differences are significant at the $P = 0.05$ level (Kruskal-Wallis test).

distributions and ecosystem carrying capacity for large grazing mammals.

This research was supported by the NSF. I thank R. S. McNaughton, N. J. Georgiadis and F. F. Banyikwa for help with the collection of samples. Laboratory analyses were performed by M. M. McNaughton. I am grateful to the Trustees and Director of Tanzania National Parks and the National Scientific Research Council of Tanzania for permission to live and do research in the Serengeti National Park and to officers of the Serengeti Wildlife Research Centre for research facilities and housing in the park.

Received 4 May; accepted 10 June 1988.

1. McNaughton, R. S. & Georgiadis, N. J. *A. Rev. Ecol. Syst.* **17**, 39–65 (1986).
2. Estes, R. D. *J. Mammal* **48**, 189–209 (1967).
3. Delany, M. J. & Hapgood, D. C. D. *Ecology of African Mammals* (Longman, London, 1979).
4. Eltringham, S. K. *The Ecology and Conservation of Large African Mammals* (Macmillan, London, 1980).
5. Vesey-Fitzgerald, D. F. *J. Mammal* **41**, 161–172 (1960).
6. Bell, R. H. V. *Scient. Am.* **224**, 1, 86–93 (1971).
7. Eisenberg, J. F. *Handbk Zool.* **8**, 1–92 (1966).
8. Jarman, P. J. *Behaviour* **48**, 215–266 (1974).
9. Sinclair, A. R. E. *J. anim. Ecol.* **54**, 899–918 (1985).
10. McNaughton, R. S. *J. Ecol. Monogr.* **55**, 259–294 (1985).
11. Leuthold, W. *African Ungulates* (Springer, New York, 1977).
12. McDowell, L. R. *Nutrition of Grazing Ruminants in Warm Climates* (Academic, New York, 1985).

13. McNaughton, R. S. *J. Science* **191**, 92–94 (1976).
14. McNaughton, R. S. *J. Ecol. Monogr.* **53**, 291–320 (1983).
15. McNaughton, R. S. *J. Am. Nat.* **124**, 863–886 (1984).
16. Serengeti Ecological Monitoring Programme, Serengeti Wildlife Research Centre, PO Box 3134, Arusha, Tanzania.
17. Thompson, L. T. & Walsh, J. N. *A Handbook of Inductively Coupled Plasma Spectrometry* (Blackie, Glasgow, 1983).
18. Munter, R. C., Halverson, T. L. & Anderson, R. D. *Comm. Soil Sci. Pl. Anal.* **15**, 1285–1322 (1984).
19. Hargrove, W. L., Ohki, K. & Wilson, D. O. *Pl. Soil.* **88**, 93–100 (1985).
20. Singh, B. R. & Misra, V. K. *Soil Sci.* **143**, 241–253 (1987).
21. Wentworth, T. R. & Davidson, E. A. *Soil Sci.* **144**, 190–202 (1987).
22. Mengel, K. & Kirkby, E. A. *Principles of Plant Nutrition* (Int. Potash Inst., Bern, 1982).
23. Barber, S. A. *Soil Nutrient Bioavailability* (Wiley, New York, 1984).
24. McDonald, P., Edwards, R. A. & Greenhalgh, J. F. D. *Animal Nutrition* (Longman, London, 1981).
25. Robbins, C. T. *Wildlife Feeding and Nutrition* (Academic, New York, 1983).
26. *The Nutrition of Herbivores* (eds Hacker, J. B. & Ternouth, J. B.) (Academic, Sydney, 1987).
27. Chapin, F. S. III *A. Rev. Ecol. Syst.* **11**, 233–260 (1980).
28. National Research Council *Nutrient Requirements of Beef Cattle* Sixth revised edn *Nutrient Requirements of Domestic Animals* Number 4 (Nat. Acad. Sci., Washington DC, 1984).
29. Weir, J. S. *Oikos* **23**, 1 (1972).
30. Risenhoover, K. L. & Peterson, R. O. *Oecologia* **71**, 121–126 (1986).
31. Denton, D. *The Hunger for Salt* (Springer, New York, 1984).
32. Jones, R. L. & Hanson, H. C. *Mineral Licks, Geophagy, and Biogeochemistry of North American Ungulates* (Iowa State Univ. Press, Ames, 1985).
33. Georgiadis, N. J. thesis, Syracuse Univ. (1987).
34. Jarrett, W. H. F. *et al. Afr. Wildl. J.* **2**, 158 (1964).
35. Albl, P., Goyazoglu, P. A. & Bezuidenhout, J. P. *Madoqua* **10**, 79 (1977).
36. Wilson, D. E. & Hirst, S. M. *Wildl. Monogr.* **4**, (1977).
37. Zumpt, I. F. & Heine, E. W. P. S. *Afr. J. Wildl. Res.* **8**, 131 (1978).
38. Mech, L. D. *Science* **198**, 320–321 (1977).
39. Rogers, L. L. *et al. J. Wildl. Mgmt.* **44**, 253–258 (1980).

Melanized dopaminergic neurons are differentially susceptible to degeneration in Parkinson's disease

Etienne Hirsch*, Ann M. Graybiel[†] & Yves A. Agid*

*INSERM U289, Laboratoire de Médecine Expérimentale, Hôpital de la Salpêtrière, 47 Bd. de l'Hôpital, 75013, Paris, France
[†]Department of Brain and Cognitive Sciences, Whitaker College, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

In idiopathic Parkinson's disease massive cell death occurs in the dopamine-containing substantia nigra^{1,24}. A link between the vulnerability of nigral neurons and the prominent pigmentation of the substantia nigra, though long suspected, has not been proved². This possibility is supported by evidence that N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) and its metabolite MPP⁺, the latter of which causes destruction of nigral neurons, bind to neuromelanin^{3,4}. We have directly tested this hypothesis by a quantitative analysis of neuromelanin-pigmented neurons in control and parkinsonian midbrains. The findings demonstrate first that the dopamine-containing cell groups of the normal human midbrain differ markedly from each other in the percentage of neuromelanin-pigmented neurons they contain. Second, the estimated cell loss in these cell groups in Parkinson's disease is directly correlated ($r = 0.97$, $P = 0.0057$) with the percentage of neuromelanin-pigmented neurons normally present in them. Third, within each cell group in the Parkinson's brains, there is greater relative sparing of non-pigmented than of neuromelanin-pigmented neurons. This evidence suggests a selective vulnerability of the neuromelanin-pigmented subpopulation of dopamine-containing mesencephalic neurons in Parkinson's disease.

Three sets of neurons were studied in autopsy material from four patients who died with a diagnosis of Parkinson's disease, and from three people without known neurological disease. These were catecholamine-containing neurons identified by tyrosine-hydroxylase (TH)⁺ immunohistochemistry, neuromelanin-pigmented (NM⁺) neurons detected in haematoxylin-eosin stained sections and acetylcholinesterase

(AChE)-positive neurons observed by acetylthiocholine iodide histochemistry. A similar analysis was made in brains from three patients who died with a diagnosis of progressive supranuclear palsy (PSP), another degenerative parkinsonian syndrome⁵. The numbers and locations of neurons of each type were recorded with a computerized plotting system for sets of serial sections through the midbrain (see Fig. 1). Maps were also made of any NM⁺, TH⁺ neurons observed in the TH-immunostained sections. Total numbers of neurons of each type were compared for each region and the percentage of NM⁺ neurons per region was calculated (see legend of Fig. 1).

In the normal midbrains (Fig. 2), TH⁺ neurons were most densely distributed within the substantia nigra pars compacta, but many also appeared in the substantia nigra pars lateralis, in tegmental cell group A8, in cell group A10 and the paramedian tegmentum, and in the central grey substance, all of which are known to contain dopaminergic neurons^{6–9}, as well as farther caudally, in the noradrenergic locus coeruleus. NM⁺ neurons, though densely concentrated in the substantia nigra pars compacta and in the locus coeruleus, were sparse in the other regions containing TH⁺ cells. They were only rarely detectable in the central grey substance.

When chartings from Parkinson's brains were compared to those from controls (Fig. 2), it was clear that many non-nigral catecholaminergic neurons were present in the Parkinson's midbrains. Loss of TH⁺ neurons was massive (77%) in the substantia nigra pars compacta but almost undetectable (3%) in the central grey substance. In cell group A8 and in the A10-paramedian tegmental region, the loss of TH⁺ neurons was intermediate (43% and 48%, respectively). A regression analysis (Fig. 3) showed a strong correlation between these estimates of catecholaminergic cell loss in the dopamine-containing cell groups in Parkinson's disease and the numbers of NM⁺ neurons in these cell groups in the control brains (Table 1). The correlation curve was linear ($P = 0.0057$), had an r value of 0.97, and showed a slope of 1.06 and an x-intercept of 4.5.

An analysis by cell group of the surviving neurons in the Parkinson's midbrains is summarized in Table 2. In the A8–A9–A10 cell complex, the relative survival of the non-pigmented subpopulation was most marked for the substantia nigra pars compacta and for tegmental cell group A8 (about three times that of the general TH⁺ population) and was moderate ($\times 1.5$) for the A10-paramedian tegmental region and substantia nigra pars lateralis. In the central grey substance, the estimated

† To whom correspondence should be addressed.

Fig. 1 Distribution of TH⁺ neurons (black dots) in a representative transverse section through the midbrain of a control specimen shown to illustrate procedures for quantitative plotting. The five catecholamine-containing regions analysed at this level are delimited by the rectilinear outlines as follows: the substantia nigra pars compacta (SNpc); the substantia nigra pars lateralis (SNpl); the medioventral part of the midbrain corresponding to the A10 cell group together with the adjoining paramedian and median midbrain tegmentum (M); the central grey substance (CGS); and the remaining part of the tegmentum dorsal to the medial lemniscus, corresponding to catecholamine-containing cell group A8 (A8). The more caudally situated locus coeruleus (not shown) was analysed only through its rostral third which extended 5,760 μ m from the rostral pole of the nucleus (identified in TH-immunostained and haematoxylin and eosin-stained sections). For each brain, x-y plots of every cell of each type analysed were obtained for regularly spaced (1,440 μ m) transverse sections using a computer-assisted microscope system that allowed each cell to be plotted once and only once. Sections were chosen from an anterior level through the subthalamic nucleus to a posterior level through the midlength of the locus coeruleus. Plots were made at $\times 100$. A grid adjustable in size to correct for relative shrinkage from one brain to another (and normalized for one control brain) was superimposed on the plots, and plots and corresponding grids were then printed. The images of serially-adjointing sections stained for AChE were projected onto each printed drawing. Landmarks such as the red nucleus (RN) and the medial lemniscus (ML), and regions of high or low AChE staining were drawn on the plots to help delimit the anatomical regions. For each region so delimited for each section, the total number of neurons of each type was counted. These values were then plotted against the cumulative distance between the sections, and the surface area under this curve gave an estimate of the total number of neurons per region from its rostral to caudal extent (see ref. 22).

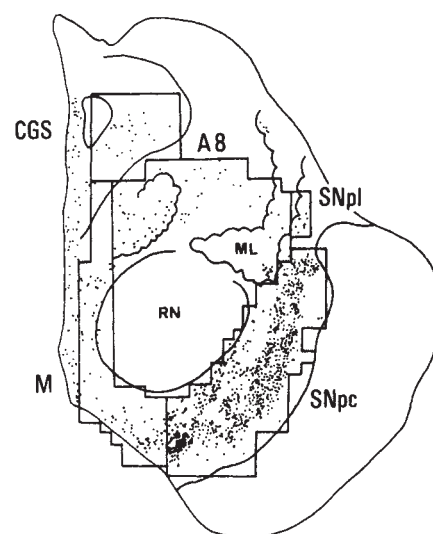


Table 1 Neuronal populations in catecholaminergic cell groups in control, parkinsonian and PSP brains

Cell group	Estimated TH ⁺ cells in controls	NM ⁺ cells in controls (%)	Decrease TH ⁺ cells in Parkinson's (%)	Decrease TH ⁺ cells in PSP (%)	TH ⁻ NM ⁺ cells in Parkinson's (%)	TH ⁻ NM ⁺ cells in PSP (%)
SNpc	213,186	84	77	84	17	64
SNpl	5,451	49	31	82	5	38
A8	38,461	52	43	65	0	0
M	32,314	51	48	76	0	19
CGS	6,836	2	3	52	0	0
lc	15,488	111	55	-7	12	3

TH⁺, TH-immunoreactive neurons; NM⁺, neuromelanin-pigmented neurons; TH⁺ NM⁺, NM⁺ TH-negative neurons. Values are % of total cells found in TH-immunostained sections of control (aged 83 ± 3.3 years, postmortem delays 12 ± 6 hours), parkinsonian (aged 65 ± 1 years, postmortem delays 27 ± 8 h) and PSP (aged 73 ± 7 years, postmortem delays 25 ± 10 h) brains. The Parkinson's and PSP cases had no other known neurological defects except for one case of Parkinson's with marked intellectual impairment. For the parkinsonian and PSP brains, the NM⁺ values are the difference between the number of TH⁻NM⁺ cells in the TH-immunostained sections and the number of NM⁺ cells counted in the haematoxylin-eosin stained sections. This avoided the possibility of including as NM⁺ neurons, dying cells or cells that do not normally express neuromelanin pigmentation (see text), and was done to obtain the most accurate estimate possible of catecholamine-containing pigmented neurons. No such TH⁻NM⁺ neurons were seen in the control brains and no correlation was found between their appearance and the postmortem delays before processing of the brains. The fact that no TH⁻NM⁺ neurons appeared in the control brains suggested that the TH-immunostaining did not yield large numbers of false negatives. In the diseased brains only if false negatives were unevenly distributed among different cell groups or among NM⁺ and NM⁻ neurons would our assessments of relative sparing have been influenced. For the PD-control comparisons, relative sparing of NM⁺ neurons in all dopaminergic cell groups was $\times 3.2$ with the correction and $\times 2.5$ uncorrected. The total number of neurons in the five dopaminergic regions studied (see Fig. 1) and in the noradrenergic locus coeruleus (lc) was estimated for each region as described in the legend of Fig. 1. The percentage of NM⁺ cells in controls was defined as the ratio of the number of NM⁺ neurons to TH⁺ neurons for each region. The percentage decrease of TH⁺ cells for each region was defined as described in the legend for Fig. 3. The percentage of TH⁻NM⁺ cells present in the parkinsonian and PSP material was defined as the ratio between the number of TH⁻NM⁺ neurons in the diseased brains, and the number of TH⁺ neurons in control brains for each region, expressed as a percentage.

survival of non-pigmented cells was 25 times that of the pigmented cells.

In the PSP midbrains, a correlation between the relative loss of TH⁺ neurons in each of these cell groups and the estimates of their normal content of NM⁺ neurons was also apparent (Fig. 3) but it was less pronounced ($P = 0.0483$). The correlation curve was linear ($r = 0.86$, slope of 1.9) with an x-intercept of 54.32. Among surviving neurons, there was marked sparing of non-pigmented neurons relative to the general TH⁺ population in all regions analysed except the A10-paramedian tegmental complex.

In contrast to these findings for the dopaminergic cell groups of the midbrain, there was no clear link between neuromelanin content and cell death in the locus coeruleus either in Parkinson's disease or in PSP (Fig. 3). Our measurements for this nucleus were limited, however, because only neurons in its anterior part could be counted (see legend of Fig. 1). We were also unable to find a consistent relationship (1) between the amount of cell death in any of the mesencephalic catecholaminergic cell groups in Parkinson's disease and PSP and the estimated total number

of TH⁺ or AChE⁺ neurons in these cell groups, or (2) between cell loss in the diseased brains and either postmortem delay before dissection or age at death.

These findings suggest that, within the entire population of dopamine-containing neurons in the midbrain, those characterized by neuromelanin pigmentation may be differentially vulnerable in Parkinson's disease and PSP. For the Parkinson's mid-brains the correlation was simple and direct: the greater the number of pigmented neurons normally present in the dopaminergic cell groups, the larger the loss of neurons in these cell groups in the diseased brains. Moreover, within each cell group, non-pigmented neurons were spared relative to the total population of TH⁺ neurons and relative to the population of pigmented neurons. For the PSP midbrains the findings also pointed to a relation between neuromelanin pigmentation and cell death in cell groups other than the A10-paramedian region, but suggested that additional cytotoxic factors were involved.

A third type of cell, not observed in the control midbrains, was identified in the Parkinson's and PSP midbrains by the presence of neuromelanin pigment in the absence of detectable

Fig. 2 Plots of TH⁺ neurons (above) and NM⁺ neurons (below) for serially-adjointing pairs of transverse sections through the nigral complex of a normal brain (control), and brains from parkinsonian patients with a diagnosis of Parkinson's disease or PSP. TH⁺ neurons were charted as described in Fig. 1 from 40 μ m transverse sections immunostained with an anti-serum against tyrosine hydroxylase (Eugene Tech.; dilution 1:120) by a double-bridge peroxidase anti-peroxidase technique²³ (reagents from Sternberger-Meyer). NM⁺ neurons were plotted from serially-adjointing sections stained by haematoxylin-eosin. Only neurons containing brown neuromelanin pigment detectable with the light microscope ($\times 100$) were plotted. The pairs of sections illustrated show the distinct difference in the distribution of TH⁺ and NM⁺ neurons in the control midbrains: the majority of TH⁺ neurons in the ventral midbrain were also NM⁺, whereas most of the TH⁺ neurons in the dorsal midbrain lacked neuromelanin. The pairs of sections from the parkinsonian brains show the sharp decrease in the number of TH⁺ cells and NM⁺ cells observed ventrally in both Parkinson's and PSP midbrains. In the dorsal neuromelanin-poor part of the midbrain, many TH⁺ neurons remain. These results were confirmed quantitatively as shown in Tables 1 and 2. RN, red nucleus; CP, cerebral peduncle.

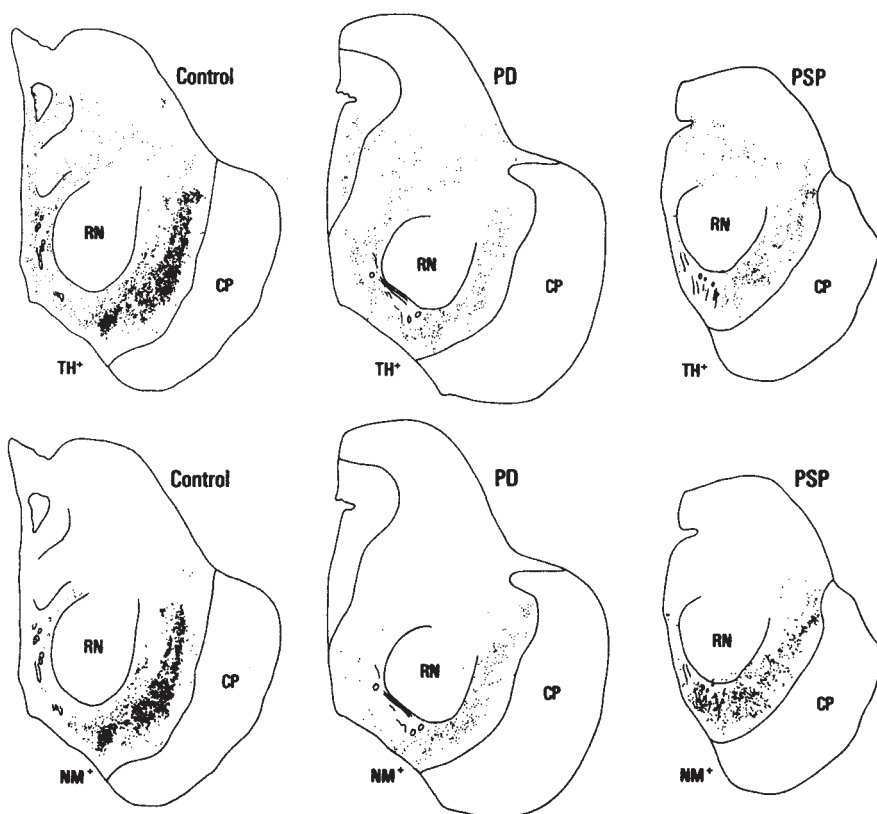
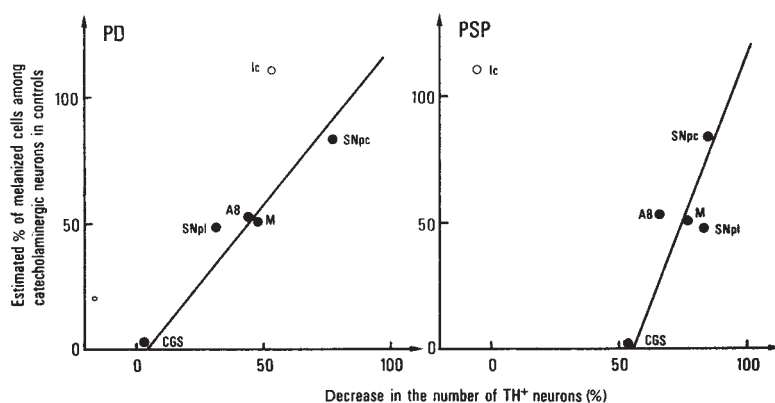


Fig. 3 Linear regression analysis illustrating correlation between the percentage of NM⁺ neurons (estimated as described in Table 1 legend) in each catecholaminergic region analysed (ordinate) and the percentage decrease in the number of TH⁺ neurons in these regions (abscissa). Left: data for Parkinson's brains. Right: data for PSP brains. The cell loss for the Parkinson's and PSP brains was estimated by the ratios between the numbers of TH⁺ neurons in each region in the diseased brains and the corresponding numbers in the control brains, expressed as percentages. The correlation curves were calculated for the five dopaminergic regions (black dots). Values for the only supposed noradrenergic region, the locus coeruleus (lc, data represented as open circles), are shown on the charts but were not included in the regression analysis. For the Parkinson's midbrains, the correlation followed a positive linear regression (slope 1.06, x intercept 4.50, $P = 0.0057$). The observed value for the locus coeruleus was more than three standard deviations away from the regression line for the dopaminergic cell groups. For the PSP midbrains the values were $r = 0.86$, slope = 1.9, x intercept = 54.32, $P = 0.483$.



TH-immunoreactivity (Table 1). These pigmented TH-negative neurons may represent dying neurons that once contained TH; alternatively, they may form a special subset of melanized neurons unique to the parkinsonian midbrains.

The distinctions we drew between the neuromelanin-positive and neuromelanin-negative cells were based on the presence or absence of visible pigmentation. It is possible that neurons without detectable pigment contained brown neuromelanin at levels below the threshold for light microscopic detection, or contained other molecular forms of neuromelanin not visible as pigment². A quantitative analysis with neuromelanin-specific staining is clearly required. Nevertheless, our findings with pigmentation as the marker for neuromelanin suggest that mechanisms underlying cell death in Parkinson's disease are related either directly to the presence of a threshold amount of neuromelanin in the affected cells or to a covariant factor, either intracellular or extracellular.

This conclusion is compatible with three main lines of evidence. First, the biochemical pathways involved in the synthesis of neuromelanin (including its derivation from dopamine) lead to the production of toxic free radicals¹⁰. Second, neuromelanin binds many drugs^{4,11} and among these is the agent MPP⁺ which is thought to produce degeneration of nigral neurons in man and in other species¹²⁻¹⁴. Third, in the normal substantia nigra, cell death appears to be related to the content of neuromelanin per cell¹⁵. Both Parkinson's and PSP are progressive diseases, so that neuromelanin-linked factors could build up with time to toxic concentrations. Preferential neurotoxicity of melanized neurons clearly does not exclude other factors in parkinsonian syndromes contributing to death of both pigmented and non-pigmented neurons¹⁶.

In our analysis, the single exception to the tight correlation between neuromelanin-pigmentation and the cell loss in catecholaminergic midbrain cell-groups was for the norepine-

Table 2 Percent survival of cell types in Parkinson's disease and PSP relative to control values

Cell group	Parkinson's disease			PSP		
	TH ⁺ cells	NM ⁺ cells	NM ⁻ cells	TH ⁺ cells	NM ⁺ cells	NM ⁻ cells
SNpc	24	18	52	16	10	43
SNpl	69	56	81	18	15	20
A8	57	29	87	35	19	53
M	52	43	61	24	35	13
CGS	97	4	99	48	0	49

TH⁺ and NM⁺ neurons were counted as described in Table 1 legend and numbers of non-pigmented TH⁺ neurons (NM⁻, were estimated as (number of TH⁺) minus (number of NM⁺). Values are expressed as percentages relative to the total population of TH⁺ neurons in control brains. There is greater survival of non-pigmented cells than pigmented cells in all five regions in Parkinson's disease. In the PSP brains, there is greater survival of non-pigmented than pigmented neurons in all of these cell groups except the A10-paramedian tegmental region (M). Abbreviations are as in Fig. 1.

phrine-containing locus coeruleus. This observation, though limited to the rostral locus coeruleus, raises the possibility that the correlation found between melanization and cell death was specific for the class of mesencephalic catecholaminergic neurons that contain dopamine.

It is important to note that the topography of cell-loss in the Parkinson's brains is consistent with cell position being an aetiological factor linked to the cellular pigmentation marker. The individual catecholamine-containing cell-groups of the mid-brain are known to have different efferent projections^{8,9,17,18}. If the initial onslaught of the disease process were at the axon terminals of these neurons rather than at their perikarya, such differential projections could be directly related to the incidence of neurotoxicity. For technical reasons we did not determine whether subdivisions within the substantia nigra pars compacta were differentially affected, but this is a reasonable expectation. In MPTP intoxication in the monkey selective sparing of parts of A8-A9-A10 cell complex and the lateral or ventral substantia nigra has been claimed^{19,20}, and in the dog, MPTP-induced degeneration of nigrostriatal fibres has been reported to start with fibres that innervate the extrastriosomal matrix²¹.

This work was funded by the Seaver Institute, the Edith C. Blum Foundation, the Tourette Syndrome Association, the McKnight Foundation, the Whitaker Foundation, the Fondation Simone et Cino del Duca, and INSERM. We thank Dr J. R. Lackner for his advice on data analysis, Dr D. Whittington for his help with the image-analysis system, and Drs J. J. Hauw, C. Duyckaerts, P. Cervera and F. Javoy-Agid for their help in obtaining the brain specimens.

Received 13 April; accepted 7 June 1988.

- Escourolle, R., De Recondo, J. & Gray, F. in *Monoamines Noyaux gris centraux et syndromes de Parkinson* (eds de Ajuriaguerra, J. & Gauthier, G.) 173-229 (Geneva, 1970).
- Marsden, C. D. *J. Neural Trans. Suppl.* **19**, 121-141 (1983).
- D'Amato, R. J., Lipman, Z. P. & Snyder, S. H. *Science* **231**, 987-989 (1986).
- D'Amato, R. J. *et al. Nature* **327**, 324-326 (1987).
- Steele, J. C., Richardson, J. C. & Olszewski, J. *Arch. Neurol.* **10**, 333-359 (1964).
- Dahlström, A. & Fuxe, K. *Acta physiol. scand.* **62**, (Suppl. 232) 1-55 (1964).
- Garver, D. L. & Sladek, J. R. *J. comp. Neurol.* **159**, 289-304 (1975).
- Graybiel, A. M. & Ragsdale, C. W. Jr in *Chemical Neuroanatomy* (ed. Emson, P. C.) 427-504 (Raven, New York, 1983).
- Lindvall, O. & Björklund, A. in *Chemical Neuroanatomy* (ed. Emson, P. C.) 229-255 (Raven, New York, 1983).
- Graham, D. G. *Molec. Pharmacol.* **14**, 633-643 (1978).
- Lindquist, N. G. *Acta radiol. Suppl.* **325**, 1-92 (1973).
- Davis, G. C. *et al. Psychiatry Res.* **1**, 249-256 (1979).
- Langston, J. W., Ballard, P., Tetrud, J. W. & Irwin, I. *Science* **219**, 979-980 (1983).
- Burns, R. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4546-4550 (1983).
- Mann, D. M. A. & Yates, P. O. *Brain* **97**, 489-498 (1974).
- Duvoisin, R. C. *Clin. Neuropharmacol.* **9**, Suppl. 1 S3-S11 (1986).
- Jimenez-Castellanos, J. & Graybiel, A. M. *Neuroscience* **23**, 223-242 (1987).

- Gerfen, C. R., Herkenham, M. & Thibault, J. J. *Neurosci.* **7**, 3915-3934 (1988).
- Deutch, A. Y. *et al. Neurosci. Lett.* **68**, 51-56 (1986).
- German, D. C., Dubach, M., Askari, S., Speciale, S. G. & Bowden, D. M. *Neuroscience* **24**, 161-174 (1988).
- Wilson, J. S., Turner, B. H., Morrow, G. D. & Hartman, P. J. *Brain Res.* **423**, 329-332 (1987).
- Hirsch, E. C., Graybiel, A. M., Duyckaerts, C. & Javoy-Agid, F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5976-5980 (1987).
- Graybiel, A. M., Hirsch, E. C. & Agid, Y. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 303-307 (1987).
- Jellinger, K. in *Recent Developments in Parkinson's Disease* (eds Fahn, S., Marsden, C. D., Jenner, P. & Teychenne, P.) 33-66 (Raven, New York, 1986).

Low retinal noise in animals with low body temperature allows high visual sensitivity

A.-C. Aho, K. Donner, C. Hydén,
L. O. Larsen* & T. Reuter

Department of Zoology, University of Helsinki, Finland
*Zoophysiological Laboratory A, August Krogh Institute,
University of Copenhagen, Denmark

The weakest pulse of light a human can detect sends about 100 photons through the pupil and produces 10-20 rhodopsin isomerizations in a small retinal area^{1,2}. It has been postulated³ that we cannot see single photons because of a retinal noise arising from randomly occurring thermal isomerizations. Direct recordings have since demonstrated the existence of electrical 'dark' rod events indistinguishable from photoisomerization signals⁴⁻⁶. Their mean rate of occurrence is roughly consistent with the 'dark light' in psychophysical threshold experiments, and their thermal parameters justify an identification with thermal isomerizations⁵. In the retina of amphibians, a small proportion of sensitive ganglion cells have a performance-limiting noise that is low enough to be well accounted for by these events⁷⁻¹⁰. Here we study the performance of dark-adapted toads and frogs and show that the performance limit of visually guided behaviour is also set by thermal isomerizations. As visual sensitivity limited by thermal events should rise when the temperature falls, poikilothermous vertebrates living at low temperatures should then reach light sensitivities unattainable by mammals and birds with optical factors equal. Comparison of different species at different temperatures shows a correlation between absolute threshold intensities and estimated thermal isomerization rates in the retina.

For the study of factors limiting the absolute sensitivity of vision, the toad *Bufo bufo* offers decisive advantages. First, it is a close relative of *Bufo marinus*, which has provided the best understanding of 'dark' rod events of any species⁵. Second, it is known to manage visually guided prey-catching at light levels so low (<10 μ lux) that a human observer cannot see the toads or their prey¹¹. Third, both the behavioural threshold of the whole animal and electrophysiological thresholds for retinal ganglion cell responses can be determined under equivalent conditions. We determined both kinds of thresholds at 15 °C and under conditions where stimulus intensities in the retina could be accurately calculated in terms of isomerizations per rhodopsin molecule per second (denoted R^*s^{-1}).

For a behavioural experiment, a fully dark-adapted toad in a plastic box was illuminated from above by a green light (525 nm) of adjustable intensity (Fig. 1). Underneath the transparent floor, white 'worm-dummies' (3 × 20 mm) moved over a black background. The dummies appeared at 12 s intervals, but the toad could not see more than one at a time. The occurrence of one or several snaps (acoustically recorded and taped) against the bottom of the plastic box was regarded as a positive response.

The results are summarized in Fig. 2a and b. The common abscissa gives stimulus intensity, that is, the isomerization rate produced by the dummy, and the black arrows mark the rate of thermal isomerizations at this temperature, $4.9 \times 10^{-12} R^*s^{-1}$.

The relative frequencies of positive snapping responses at various intensities are given by the filled circles in Fig. 2a. For comparison, Fig. 2b displays thresholds (small dots) and the intensity ranges in which responses could be obtained (horizontal lines) for 18 well characterized retinal ganglion cells in eyecup preparations stimulated with rectangles of green light, equivalent to the retinal image of the dummy in the behavioural experiments. Figure 2c exemplifies the response variability of one of the ganglion cells when stimulated ten times with three fixed intensities, one just below and two just above the 50% response threshold.

All toads responded promptly in the intensity range from $3 \times 10^{-11} \text{ R}^* \text{ s}^{-1}$ upwards, and the same applied to 14 out of 18 ganglion cells. But between 3×10^{-11} and $3 \times 10^{-12} \text{ R}^* \text{ s}^{-1}$, both ganglion cell and snapping activity decreased sharply. The 50% response threshold of the most sensitive ganglion cell was $3.1 \times 10^{-12} \text{ R}^* \text{ s}^{-1}$, and that was also the lowest intensity for snapping that was obviously light-dependent. One of the four toads responding at that intensity did snap once even at $6 \times 10^{-13} \text{ R}^* \text{ s}^{-1}$, and also snapped once in complete darkness. As no other toad ever responded at lower intensities, the light-dependence of the snapping at $3.1 \times 10^{-12} \text{ R}^* \text{ s}^{-1}$ (4 toads snapping out of 18 tested) is statistically significant at $P < 0.01$.

It appears remarkable that the toads can detect an object that adds only $3.1 \times 10^{-12} \text{ R}^* \text{ s}^{-1}$ to the ongoing rate of $4.9 \times 10^{-12} \text{ R}^* \text{ s}^{-1}$ from thermal isomerizations. But the limit to detection is set by random (Poisson) fluctuations in the number of isomerizations summed by a detector within consecutive counting intervals (summation times), not by rates as such. To evaluate whether thermal events really limit detection, it is necessary to calculate signal-to-noise ratios based on the relevant numbers¹². For ganglion cells, summation areas and times are easily defined and the calculation is straightforward⁷⁻¹⁰. The most sensitive cell in Fig. 2b had a summation time of 1.9 s and collected signals from 440 rods (each containing about 3.25×10^9 molecules of rhodopsin). Thus at threshold it collected a mean number of 8.4 photoisomerizations, together with 13.3 thermal isomerizations, giving the upper limit $8.4/(8.4 + 13.3)^{1/2} = 1.8$ for the signal-to-noise ratio of the response. This indicates a low reliability of detection¹³.

The brain of the toad may improve on the poor signal-to-noise ratio by summing signals from several sensitive ganglion cells. Uncertainties regarding the extent of central summation, sampling of the visual field, the role of binocular vision and so forth, preclude a strictly quantitative treatment, but it is instructive to consider the border conditions. The dummy image covered $\sim 4,500$ rods; within the summation time 1.9 s used above (the average among the 18 ganglion cells was in fact somewhat shorter at 1.5 s), the mean number of photoisomerizations collected over the whole image would be 86. During the same summation time the mean number of thermal isomerizations within the same area is 136. But the brain has no means of restricting its attention to that particular area and must sample the entire retina, which is 130 times larger.

Truly noise-limited performance must deteriorate if the noise increases, and this yields an interesting and directly testable prediction. The rate of thermal isomerizations grows about 4-fold for every 10°C temperature rise⁵. If this noise constitutes a limiting factor, then the lowest possible threshold (in terms of $\text{R}^* \text{ s}^{-1}$) of animals active at high body temperatures should be higher than that of our toads at 15°C (provided that temporal summation is not significantly larger in warmer animals).

To test this idea, we have determined absolute thresholds in man, and frogs at different body temperatures. Again, we expressed the thresholds in terms of photoisomerizations per rhodopsin molecule per second, thus eliminating all differences in optics anatomy of the rods and compared them with the estimated rates for thermal isomerizations. The human thresholds were determined with the same prey-dummies in the same apparatus as was used for the toads; the frequency-of-seeing

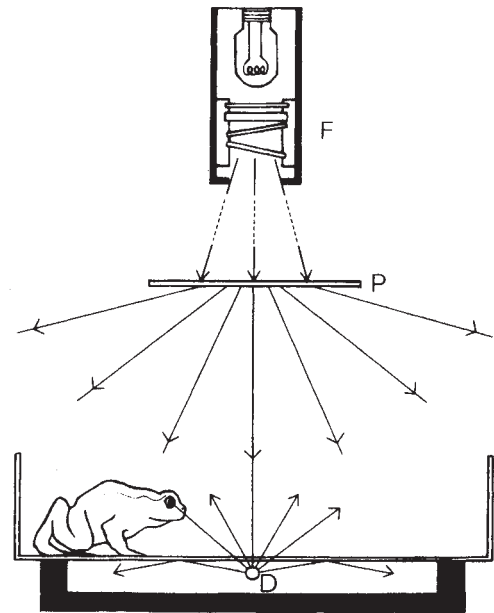


Fig. 1 Schematic presentation of the experimental set-up in the snapping experiments¹¹. The apparatus was located in a light-tight black-painted room. The well shielded lamp was connected to a stabilized current source. *F* represents heat-absorbing filters, neutral density filters and the 525 nm interference filter, and *P*, the plastic diffuser screen on which the lamp formed a luminous 39 cm² disc 35 cm above the toad. Direct retinal illumination from this 'moon' would be 300 times more intense than that from the dummy, but control experiments in which the source was screened from direct view indicated that its potential visibility had no clear negative effect on snapping sensitivity. *D* is a cross-section through a white worm-dummy (3 × 20 mm, attached to a transparent string and moved at a speed of 8.3 mm s⁻¹ and a frequency of one per 12 s). The toad could not see more than one dummy at a time. Optical calculations were checked by direct measurements of reflections, transmissions and absolute intensities using a reflection spectrophotometer¹⁴ and a radiometer calibrated against rhodopsin extracts⁸. The distance between the eye of the toad and the dummy could not be accurately controlled, but was estimated to be 50 mm (ref. 15). Variation in this distance affects the size of the retinal image but not its intensity. For 50 mm distance and 4.2 mm posterior focal length of the eye¹⁶, the image of the dummy occupied 0.39 mm² of the retina and 4,500 red rods. The entrance pupil (7.1 ± 0.3 mm²) was determined by flash photography face-on; 9% of the incoming light was estimated to be reflected from the curved cornea⁹. No corrections were made for possible light losses in the lens, vitreous and neural retina. 34% of 525 nm photons incident on the *Bufo bufo* retina produce isomerizations in red rods (for methods, see ref. 8). 18 carefully selected¹¹ snapping male toads (*Bufo bufo*, eye diam. 6.0–6.5 mm) from southern Finland were used. They were kept at 12–16 °C, with access to water and dark and moist shelters. All prey-catching experiments were carried out in July and August, noon to 1800 h, at $\sim 15^\circ \text{C}$. The toads were always allowed one day of rest between two tests. They were only fed just after each test. Increased reluctance to snap at dummies was observed after two months.

function of one subject is plotted as open circles in Fig. 2a. For frogs, we determined the absolute threshold for the phototactic jumping behaviour at three different temperatures⁹. These results are summarized in Fig. 3 as a log-log plot of threshold intensities against rates of thermal isomerizations.

A monotone relation is found between threshold intensity and rate of thermal isomerizations. This is consistent with the hypothesis that thermal isomerization of rhodopsin sets the ultimate limits on threshold intensity for the visual detection task. This could be ecologically significant, for example for vision in the cold and dark environment inhabited by deep-sea

Fig. 2 Behavioural (*a*) and electrophysiological (*b, c*) determinations of absolute sensitivity at $\sim 15^\circ\text{C}$. The abscissa in (*a*) and (*b*) gives the intensity of illumination in the retina as numbers of photoisomerizations per rhodopsin molecule per second ($R^* s^{-1}$; logarithmic scale). Retinal stimulus in all cases (except for human psychophysics) was 0.39 mm^2 rectangle of light covering 4,500 rods. The thermal isomerization rate ($4.9 \times 10^{-12} R^* s^{-1}$) is marked by black arrows on the abscissa and has been obtained from the rate of 'dark' rod events found in *Bufo marinus*⁵, and corrected for temperature. *a*, Snapping behaviour of *Bufo bufo*: each filled circle represents an experiment with nine toads; a response frequency of 100 on the ordinate means that all nine (9/9) snapped within 120 s after being placed in the apparatus (Fig. 1). Human psychophysics (open circles): frequency-of-seeing function determined in the same apparatus. The function is placed so that the abscissa directly gives isomerization rates in the human retina. This entails a slight shift in relation to the toad function, so that the 12-fold difference at the 50% response level corresponds to an 8-fold difference in the corresponding light intensities illuminating the dummy. The subject (A.-C.A.) aged 24 years, was given 60 min to dark-adapt; the cornea-dummy distance was 160 mm, her head was fixed but the eyes moved freely. The light source (P in Fig. 1) was shielded from view and the pupil area was 37 mm^2 . 100 tests were performed in semi-random order: 20 at each of four intensities and 20 in darkness. The criterion for seeing corresponded to a 'glimpse of something' during a 25 s concentrated binocular search while two dummies passed, with a 30–60 s pause between searches. For the human subject, the test size corresponded to 1.1×7.2 degrees of visual angle (0.61 mm^2 on the retina) and was thus sufficient for full scotopic spatial summation. Assuming that 22% of 525 nm photons entering at the cornea produce isomerizations², and taking full dark-adapted spatio-temporal summation as 0.185 s deg^2 (ref. 17), the 50% threshold corresponds to 14 isomerizations in one retina, in good agreement with thresholds obtained with small flashed spot stimuli^{1,2}. For calculation of human intensities in terms of $R^* s^{-1}$, see legend to Fig. 3. The two dotted curves are cumulative Poisson curves with thresholds and noise values determined for maximum-likelihood fit to the points.^{7,8,10} *b* and *c*: extracellularly recorded ganglion cell spike responses from isolated eyecups⁹. *b*, Horizontal lines give the ranges of stimulus intensities for which each of 18 cells responded to the rectangle; the small terminating dot in each case marks the intensity to which the cell responded on 50% of the trials. This 50% threshold turned out to be independent of whether the rectangle was moved across the receptive field at 'dummy-speed', or whether it was presented as an on-step of light. *c*, Response variability around threshold of the cell marked with a star in *b*. Here step stimulation was used to obtain precise time-courses. The three panels show responses to ten presentations of one fixed intensity each (log intensity in $R^* s^{-1}$ indicated in the upper left-hand corner). Each row of spikes shows one discharge on a time abscissa; stimulus onset at the left edge of the panel. The cell is seen to respond to all presentations of the highest intensity, but only twice to the lowest intensity, which is thus sub-threshold. This medium-sensitive cell had a very low maintained activity, giving only 5 'spontaneous' spikes during 1 h of recording.

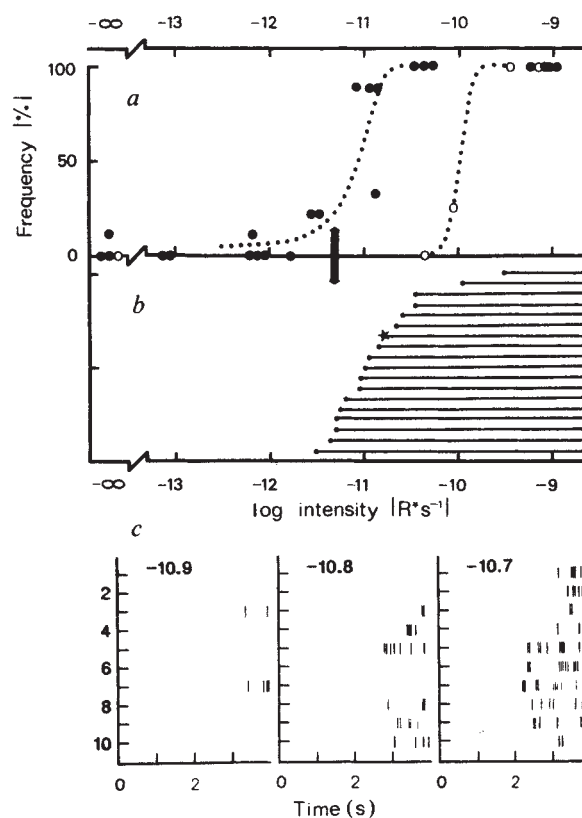
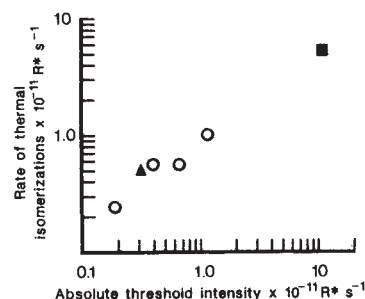


Fig. 3 Correlation between rates of thermal rhodopsin isomerizations (ordinate) and absolute threshold intensities (abscissa), expressed as rates of isomerizations per rhodopsin molecule in the retina of the toad, *Bufo bufo* ($\blacktriangle$), the frog, *Rana temporaria* ($\circ$), and man ($\blacksquare$). The toad (15°C) and human (37°C) results are taken from Fig. 2*a*; the frog results were obtained at temperatures (from left to right) 10°C , 16°C , 16°C and 20°C . For methods, see ref. 9. For proper appreciation of the differences, note that the sensitivity of frog rods (response amplitude/photoisomerization) is in fact lower at 10°C than at 16 or 20°C (ref. 18). For the amphibians, threshold is the lowest intensity for which the light-dependence of the behaviour was statistically significant ($P < 0.01$). In the phototaxis experiments with *Rana*, the spacing of trial intensities was ~ 0.3 log units. The human threshold (50% responses) is in good agreement with the generally accepted value of 10–20 quanta used², if full dark-adapted spatio-temporal summation is assumed¹⁷. Amphibian rates of thermal isomerizations are taken from ref. 5 and corrected for temperature; human rate of thermal isomerizations is from ref. 6. Human photoisomerizations per rhodopsin molecule have been calculated assuming there are 1.2×10^8 molecules per rod⁶ and a peak rod density of 160,000 per mm^2 retina¹⁹.



fishes. It also emphasizes the functional importance of any anatomical structures that increase the numbers of photons absorbed without increasing the amount of absorbing pigment (the number of noise-producing rhodopsin molecules). Well known examples of such structures are reflecting tapeta and light-funneling inner segments of photoreceptors.

This work was supported by the Academy of Finland.

1. Hecht, S., Schlaer, S. & Pirenne, M. H. *J. gen. Physiol.* **25**, 819–840 (1942).
2. Barlow, H. B. in *Vertebrate Photoreception* 337–358 (Academic, London, 1977).
3. Barlow, H. B. *J. opt. Soc. Am.* **46**, 634–639 (1956).
4. Ashmore, J. F. & Falk, G. *Nature* **270**, 69–71 (1977).
5. Baylor, D. A., Matthews, G. & Yau, K.-W. *J. Physiol.* **309**, 591–621 (1980).
6. Baylor, D. A., Nunn, B. J. & Schnapf, J. F. *J. Physiol.* **357**, 575–607 (1984).
7. Reuter, T., Donner, K. & Copenhagen, D. R. *Neurosci. Res. suppl.* **4**, 163–180 (1986).
8. Copenhagen, D. R., Donner, K. & Reuter, T. *J. Physiol.* **393**, 667–680 (1987).
9. Aho, A.-C. *et al. J. opt. Soc. Am. A* **4**, 2321–2329 (1987).
10. Donner, K. *Physica Scripta* (in the press).
11. Larsen, L. O. & Pedersen, J. N. *Amphibia-Reptilia* **2**, 321–327 (1982).
12. Barlow, H. B. in *Photophysiology* Vol. 2, 163–202 (Academic, New York, 1964).
13. Rose, A. *J. opt. Soc. Am.* **38**, 196–208 (1948).
14. Maximov, V. V., Orlov, O. Yu & Reuter, T. *Vision Res.* **25**, 1037–1049 (1985).
15. Burghagen, H. & Ewert, J.-P. *J. comp. Physiol. A* **152**, 241–249 (1983).
16. Du Pont, J. S. & de Groot, P. J. *Vision Res.* **16**, 803–810 (1974).
17. Barlow, H. B. *J. Physiol.* **141**, 337–350 (1958).
18. Donner, K., Hemilä, S. & Koskelainen, A. *Acta physiol. scand.* (in the press).
19. Østerberg, G. *Acta Ophthalm. suppl.* **6**, 1–103 (1935).

Received 5 April; accepted 16 June 1988.

Sensory transmitters regulate intracellular calcium in dorsal horn neurons

M. D. Womack*†, A. B. MacDermott†‡
& T. M. Jessell†§

* Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA

† Howard Hughes Medical Institute, Center for Neurobiology and Behavior,

‡ Department of Physiology and Cellular Biophysics,

§ Department of Biochemistry and Molecular Biophysics, Columbia University, College of Physicians and Surgeons, New York, New York 10032, USA

Primary afferent terminals in the dorsal horn of the spinal cord release excitatory amino acid and peptide transmitters that initiate the central processing of nociceptive information¹. The postsynaptic actions of amino acid transmitters on spinal neurons have been well characterized², but the cellular basis of peptide actions remains unclear³. Substance P is the best characterized of the peptides present in sensory neurons and has been shown to depolarize dorsal horn neurons and to facilitate nociceptive reflexes⁴⁻⁷. To determine the mechanisms by which substance P contributes to afferent synaptic transmission, we have monitored the levels of intracellular calcium in single isolated rat dorsal horn neurons and report that substance P can produce a prolonged elevation in calcium concentration by mobilizing its release from intracellular stores. This elevation may contribute to the long-term changes in the excitable properties of dorsal horn neurons that occur following afferent fibre stimulation⁶. We have also found that L-glutamate elevates intracellular calcium in substance P-sensitive dorsal horn neurons by increasing calcium influx^{8,9}. These results provide a direct demonstration of intracellular calcium changes in response to neuropeptides in mammalian central neurons. They also indicate that there is convergent regulation of intracellular calcium in dorsal horn neurons by two different classes of sensory transmitters that are co-released from the same afferent terminals.

The intracellular calcium concentration ($[Ca^{2+}]_i$) of single isolated postnatal rat dorsal horn neurons was monitored with the Ca^{2+} indicator fura-2. Neurons were examined 2–8 hours after isolation, before the formation of synaptic connections, thus permitting analysis of the direct effects of substance P. Of the 544 dorsal horn neurons tested, 141 (26%) responded to substance P with an increase in $[Ca^{2+}]_i$ (Fig. 1). The peak $[Ca^{2+}]_i$ in response to substance P (10–100 nM) varied between 100 nM and 2.7 μ M (mean $721 \pm s.e.m.$ 83 nM, $n = 39$ neurons). The mean resting $[Ca^{2+}]_i$ ($42 \pm s.e.m.$ 4 nM, $n = 50$ neurons) in substance P-responsive and nonresponsive neurons was not significantly different. The resting Ca^{2+} level in dorsal horn neurons is comparable to values obtained with other classes of neurons from the central nervous system^{9,10}, but is lower than values reported for certain non-neural cells¹¹. In some dorsal horn neurons, resting $[Ca^{2+}]_i$ was determined to be 10 nM or less, which is probably an underestimate of the true $[Ca^{2+}]_i$. These low values may result from incomplete de-esterification of the acetoxymethyl ester form of fura-2¹² or from sequestration of the dye in noncytosolic compartments in the cell¹³.

The minimum effective concentration of substance P varied between 1 and 10 nM (Fig. 1), however the peak $[Ca^{2+}]_i$ did not increase markedly with increases in substance P concentration beyond 10 nM (not shown). To determine the subclass of tachykinin receptors involved in the substance P-mediated elevation of $[Ca^{2+}]_i$, we compared the effects of substance P, which acts primarily at the NK1 class of tachykinin receptors¹⁴, with that of neurokinin A (substance K) which has a markedly higher affinity than substance P for NK2 and NK3 receptors¹⁴. Activation of both NK1 and NK2 receptors has been reported to enhance intracellular Ca^{2+} levels in non-neural cells¹⁵⁻¹⁷. Dorsal

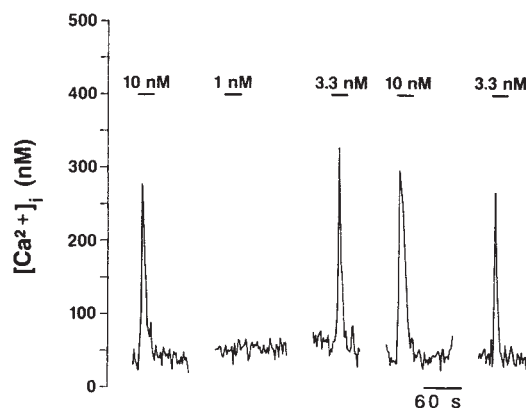


Fig. 1 Increases in $[Ca^{2+}]_i$ in a single dorsal horn neuron induced by substance P. Substance P (at indicated concentrations) was applied from a pressure pipette (1.5 p.s.i.) over the period indicated by the horizontal bars. Gaps in the record represent periods during which the signal to the photomultiplier was blocked in order to change pipettes. Approximately 5 min elapsed between drug applications.

Methods. Acutely dissociated dorsal horn cells were obtained from rats at postnatal days 0–3. Cultured neurons were isolated from E18 embryos. Cell suspensions from both ages were prepared from transverse spinal cord slices (500 μ m) cut on a McIlwain tissue chopper and dissociated by incubation in 0.064% trypsin (Sigma, type III) for 35 min³⁰. Single cells were plated in supplemented Ham's F12 medium³¹ on a monolayer of rat cortical astrocytes grown on glass coverslips. For recording, cells were washed with BSS (142 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, supplemented with 10 mM HEPES, 8 mg ml⁻¹ glucose and 0.4 mg ml⁻¹ BSA, pH 7.2) and incubated for 10 min at 37 °C in BSS containing 2 μ M fura-2-AM, then washed and incubated for 10 min at 37 °C in BSS. Coverslips containing dye-loaded cells were mounted into a lucite holder that formed a chamber through which BSS (22 °C) was perfused at 10 ml min⁻¹. Neurons were identified by their morphology and by expression of neuronal surface antigens L1 and NCAM (M.D.W. *et al.* in preparation). Recording of fluorescence from cell bodies of single neurons was carried out on a Zeiss ICM-405 microscope at alternating (0.5 Hz) excitation wavelengths of 340 and 380 nm. Fluorescence signals were recorded with a DCP-2 photometer (Zeiss). Fluorescence ratios were obtained by subtracting the background fluorescence recorded from a blank area on the glass coverslip and converted to $[Ca^{2+}]_i$ using the equation $[Ca^{2+}]_i = A(R - R_{min}) / (R_{max} - R)$ (ref. 32). The value of A was determined using solutions of fura-2 (1 μ M in 125 mM KCl, 20 mM NaCl, 10 mM HEPES, 1 mM Mg^{2+} , pH 7.0, 22 °C) in which Ca^{2+} was buffered to known concentrations with 10 mM EGTA.

horn neurons that respond to substance P at concentrations of 1–10 nM were responsive to neurokinin A only at concentrations >100 nM ($n = 14$ neurons). Moreover, dorsal horn neurons originally identified by their response to neurokinin A were sensitive to lower concentrations of substance P ($n = 3$). The tachykinin receptors that mediate the substance P-induced elevation in $[Ca^{2+}]_i$ therefore appear to be of the NK1 class. These observations suggest either that the NK2 class of receptors is not expressed by dorsal horn neurons, or that NK2 receptors are coupled to a transduction mechanism that does not result in an increase in $[Ca^{2+}]_i$. No change in $[Ca^{2+}]_i$ was detected in response to application of neurokinin B (100 nM), a preferential NK3 receptor agonist ($n = 20$ neurons).

The elevation in $[Ca^{2+}]_i$ in response to substance P could result from mobilization of Ca^{2+} from intracellular stores or from Ca^{2+} entry. Superfusion of substance P-sensitive dorsal horn neurons with a nominally Ca^{2+} -free medium (Fig. 2a, b) did not significantly reduce (by less than 10%) the amplitude of the response to substance P in 56% (22/39) of neurons examined. The persistence of the response to substance P did not appear to result from the inadequate reduction of extracellular Ca^{2+} . Increases in $[Ca^{2+}]_i$ in response to L-glutamate and

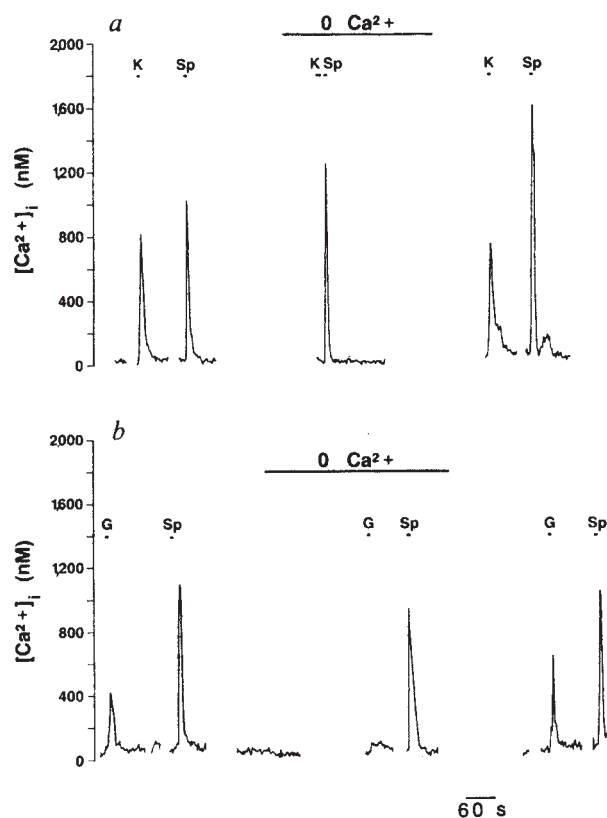


Fig. 2 Increases in $[Ca^{2+}]_i$ evoked by substance P are not affected by lowering the extracellular $[Ca^{2+}]$. *a*, The first recording was obtained from an acutely dissociated neuron which responded to pressure application (5 s) of both 100 nM substance P (Sp) and high potassium buffer (K) (21 mM NaCl, 126 mM KCl, 1 mM $MgCl_2$, 10 mM HEPES, 8 mg ml^{-1} glucose and 0.4 mg ml^{-1} BSA $\pm$ 2 mM $CaCl_2$). The middle trace was recorded 3 min after superfusion of Ca^{2+} -free medium (BSS containing no added $CaCl_2$). In Ca^{2+} -free medium the cell did not respond to a ten-second application of K^+ , however subsequent application of substance P produced an increase in $[Ca^{2+}]_i$. When the external Ca^{2+} concentration was restored to 2 mM, the response to K^+ recovered. Approximately 10 min elapsed between applications of substance P. *b*, Recordings obtained from a neuron which responded to both 100 nM substance P and 20 μ M L-glutamate (G). The increase in $[Ca^{2+}]_i$ elicited by L-glutamate was greatly decreased during superfusion with medium containing no added Ca^{2+} , whereas the response to substance P was unaffected.

elevated levels of potassium (K^+), which have been shown to promote Ca^{2+} entry^{8,9}, were virtually abolished by removal of extracellular Ca^{2+} (Fig. 2*a, b*). Thus substance P-induced increases in $[Ca^{2+}]_i$ can be mediated by the mobilization of intracellular Ca^{2+} stores. Dorsal horn neurons that responded to substance P in the presence but not in the absence of extracellular Ca^{2+} may represent a distinct population of neurons in which substance P does promote Ca^{2+} entry. Alternatively, intracellular Ca^{2+} stores in these neurons may have been depleted by exposure to low extracellular Ca^{2+} prior to application of substance P.

The duration of changes in $[Ca^{2+}]_i$ was then examined. Brief (1–5 s) applications of substance P raised $[Ca^{2+}]_i$ in dorsal horn neurons for periods of 10–240 s. In neurons in which substance P evoked Ca^{2+} release from intracellular stores, application of substance P for periods of up to 2 min did not prolong the duration of elevated $[Ca^{2+}]_i$ (Fig. 1). The time course of the substance P-induced elevation in $[Ca^{2+}]_i$ in these neurons, therefore, appears to be independent of the duration of peptide application. A 1–10 min interval was required before neurons responded to a second application of substance P. The transient nature of the $[Ca^{2+}]_i$ increase in response to maintained peptide

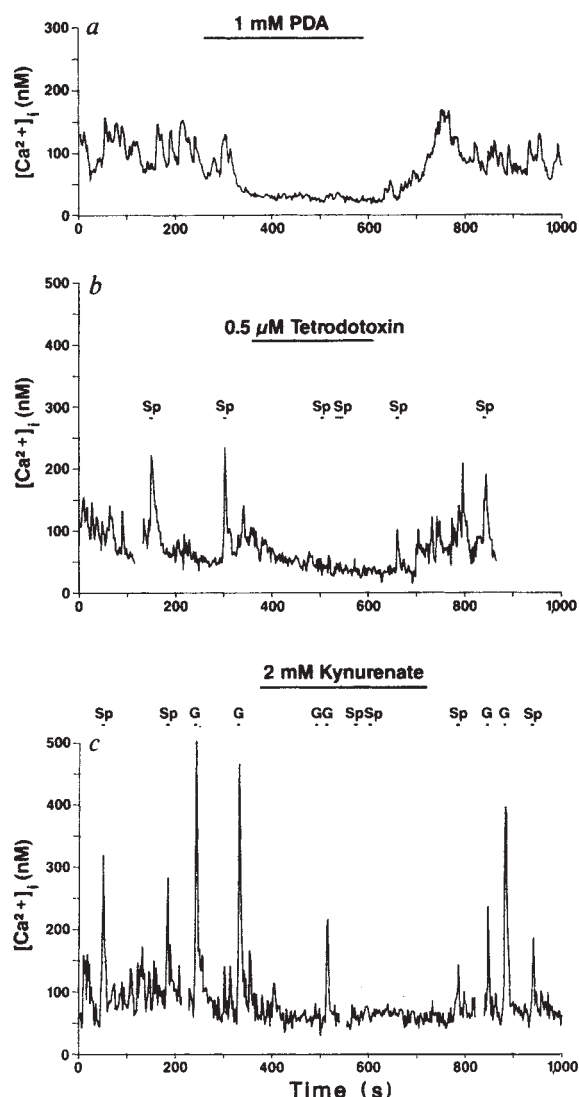


Fig. 3 Substance P increases $[Ca^{2+}]_i$ in cultured dorsal horn neurons by evoking release of excitatory amino acid transmitters. *a*, Recording of $[Ca^{2+}]_i$ in a neuron maintained in culture for 8 days in which spontaneous Ca^{2+} fluctuations were detected. Superfusion of 1 mM $\pm$ cis 2,3-piperidine dicarboxylic acid (PDA), an excitatory amino acid antagonist, blocked the $[Ca^{2+}]_i$ fluctuations. In 7/7 cells, fluctuations were blocked by superfusion with tetrodotoxin (data not shown) and in 11/12 neurons by superfusion with PDA (1 mM) or kynurenat (2 mM). *b*, Increases in $[Ca^{2+}]_i$ in response to pressure application of 100 nM substance P (Sp) are blocked by superfusion of 0.5 μ M tetrodotoxin. Responses to substance P were blocked by tetrodotoxin in 4/7 cultured dorsal horn neurons. *c*, Responses to both 100 nM substance P (5 s) and 20 μ M L-glutamate (G, 1 s) are blocked by superfusion with 2 mM kynurenat. A longer application of L-glutamate (5 s) during kynurenat perfusion produces a small increase in $[Ca^{2+}]_i$, possibly because kynurenat is washed away from the cell. Similar results were obtained in 5/12 neurons. Recovery of responses to substance P and L-glutamate occur after kynurenat is washed out.

application may reflect the desensitization of substance P receptors, the inactivation of other cellular components involved in the transduction of substance P actions, or the depletion of intracellular Ca^{2+} stores.

We found that 83% of substance P-responsive dorsal horn neurons (10/12) also exhibited an increase in $[Ca^{2+}]_i$ in response to L-glutamate (20 μ M). L-glutamate evoked an increase in $[Ca^{2+}]_i$ in a similar proportion (82%, 28/34) of substance P-insensitive neurons. The peak $[Ca^{2+}]_i$ in response to L-glutamate ranged from 200 nM to 1.7 μ M (mean $660 \pm$ s.e.m. 100 nM, $n = 15$ neurons). The actions of L-glutamate appear to be mediated

via both NMDA and non-NMDA receptors since most dorsal horn neurons responded to the agonists kainate or quisqualate while about 30% were sensitive to NMDA (not shown). These observations indicate that sensory transmitters co-stored in terminals of small diameter primary afferents¹⁸ can interact directly with the same subset of postsynaptic dorsal horn neurons.

In the dorsal horn, substance P may also elevate neuronal $[Ca^{2+}]_i$ indirectly by evoking the release of acidic amino acid or other excitatory transmitters from local interneurons. To examine this possibility, we maintained embryonic rat dorsal horn neurons in dissociated cell culture for 3–8 days. Many cultured dorsal horn neurons exhibited spontaneous fluctuations in $[Ca^{2+}]_i$ that were blocked by superfusion with tetrodotoxin (1 μ M) (data not shown) and with the excitatory amino acid antagonists $\pm$ cis-2,3-piperidine dicarboxylic acid (1 mM) (Fig. 3a) or kynurenate (2 mM). These fluctuations in $[Ca^{2+}]_i$ therefore appear to be caused by synaptic activation that is mediated by the release of excitatory amino acid transmitters from nearby neurons. Substance P (100 nM) increased $[Ca^{2+}]_i$ in 68/200 (34%) cultured dorsal horn neurons. In about 50% of these neurons, the increase in $[Ca^{2+}]_i$ induced by substance P was blocked by tetrodotoxin (1 μ M) (Fig. 3b) and by kynurenate (2 mM) (Fig. 3c). Substance P can therefore enhance $[Ca^{2+}]_i$ *in vitro* via an indirect mechanism that involves release of amino acid transmitters from dorsal horn neurons. Substance P may exert a similar function *in vivo*.

Our results establish that substance P and excitatory amino acid receptors are co-expressed on single dorsal horn neurons. Both classes of transmitters appear to regulate $[Ca^{2+}]_i$ in these neurons. Substance P (100 nM) increased $[Ca^{2+}]_i$ in 68/200 $[Ca^{2+}]_i$ in many dorsal horn neurons by mobilizing intracellular Ca^{2+} stores. The intracellular pathways linking substance P receptor activation to Ca^{2+} release have not been defined, but, by analogy with non-neuronal systems¹⁹, may involve the generation of inositol phosphate intermediates²⁰. The dorsal horn has, however, been reported to lack inositol trisphosphate receptors²¹, raising the possibility that other second messenger systems may contribute to the substance P-induced rise in $[Ca^{2+}]_i$. In contrast, the increases in $[Ca^{2+}]_i$ following application of L-glutamate appear to result from Ca^{2+} entry, either through voltage-gated Ca^{2+} channels activated as a consequence of depolarization or via NMDA-gated channels^{8,9,22}. The convergent regulation of $[Ca^{2+}]_i$ may represent one cellular mechanism by which interactions between peptide and amino acid sensory transmitters are achieved.

Increases in $[Ca^{2+}]_i$ evoked by substance P may contribute to the regulation of membrane conductance and to longer-term changes in the biochemical and excitable properties of dorsal horn neurons. Calcium-activated potassium and chloride currents and Ca^{2+} -dependent spike-afterpotentials have been detected in dorsal horn neurons^{23–25} and the peak $[Ca^{2+}]_i$ evoked by both substance P and L-glutamate is in the range that has been reported to activate Ca^{2+} -dependent currents in spinal neurons²⁶. The induction of *c-fos* expression observed in response to stimulation of high threshold primary afferents²⁷ provides one potential mechanism by which a substance P-induced elevation in $[Ca^{2+}]_i$ could contribute to longer-term changes in the functional properties of dorsal horn neurons^{28,29}.

We thank R. N. McBurney, L. Role, S. Siegelbaum, and M. Tessier-Lavigne for critical reading of an earlier version of the manuscript, M. W. Jessell for computer programming, A. Gault for help with cell culture and R. Lenertz for preparing the manuscript. This work was supported, in part, by grants from the National Multiple Sclerosis Society, the McKnight Foundation and the NIH. T.M.J. is an investigator of the Howard Hughes Medical Institute.

3. Jessell, T. M. & Womack, M. D. *Trends Neurosci.* **8**, 43–45 (1985).
4. Murase, K. & Randic, M. *J. Physiol., Lond.* **346**, 203–217 (1984).
5. Nowak, L. M. & MacDonald, R. L. *J. Neurosci.* **2**, 1119–1128 (1982).
6. Woolf, C. & Wiesenfeld-Hallin, Z. *Neurosci. Lett.* **66**, 226–230 (1986).
7. Criddle, R. A. & Henry, J. L. *Brain Res.* **381**, 93–99 (1986).
8. MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J. & Barker, J. L. *Nature* **321**, 519–522 (1986).
9. Murphy, S. N., Thayer, S. A. & Miller, R. J. *J. Neurosci.* **7**, 4145–4158 (1987).
10. Connor, J. A. & Hockberger, P. E. *J. Neurosci.* **7**, 1384–1400 (1987).
11. Tsien, R. Y. & Rink, T. J. *Curr. Meth. cell. Neurobiol.* **3**, 249–312 (1983).
12. Scanlon, M., Williams, D. A. & Fay, F. S. *J. Biol. Chem.* **262**, 6308–6312 (1987).
13. Almers, W. & Neher, E. *FEBS Lett.* **192**, 1–18 (1985).
14. Lee, C.-M., Campbell, N. J., Williams, B. J. & Iversen, L. L. *Eur. J. Pharmac.* **130**, 209–217 (1986).
15. Womack, M. D., Hanley, M. R. & Jessell, T. M. *J. Neurosci.* **5**, 3370–3378 (1986).
16. Payan, D. G. *Biochem. biophys. Res. Commun.* **130**, 104–109 (1985).
17. Masu, Y. *et al.* *Nature* **329**, 834–836 (1987).
18. DeBiasi, S., VanEyck, S. L., Petruz, P. & Rustioni, A. *Am. Soc. Neurosci. Abstr.* **13**, 1563 (1987).
19. Hanley, M. R., Lee, C. M., Jones, L. M. & Michell, R. H. *Molec. Pharmac.* **18**, 78–83 (1980).
20. Mantyh, P. W., Pinnock, R. D., Downes, C. P., Goedert, M. & Hunt, S. P. *Nature* **309**, 795–797 (1984).
21. Worley, P. F., Baraban, J. M., Colvin, J. S. & Snyder, S. H. *Nature* **325**, 159–161 (1987).
22. Mayer, M. L., MacDermott, A. B., Westbrook, G. L., Smith, S. J. & Barker, J. L. *J. Neurosci.* **7**, 3230–3244 (1987).
23. Schwandt, P. C. & Crill, W. E. *J. Neurophysiol.* **46**, 1–16 (1981).
24. Owen, D. G., Segal, M. & Barker, J. L. *Nature* **311**, 567–570 (1984).
25. Yoshimura, M., Polosa, C. & Nishi, S. *J. Neurophysiology* **55**, 1234–1246 (1986).
26. Blair, L. E. & Dionne, V. E. *Nature* **315**, 329–331 (1985).
27. Hunt, S. P., Pini, A. & Evan, G. *Nature* **328**, 632–634 (1987).
28. Morgan, J. I. & Curran, T. *Nature* **322**, 552–555 (1986).
29. Wall, P. D. & Fitzgerald, M. *CIBA Fdn Symp.* **91**, 249–260 (1982).
30. Kay, A. R. & Wong, R. K. S. *J. Neurosci. Meth.* **16**, 227–238 (1986).
31. Romijn, H. J., Habets, A. M. M. C., Mud, M. T. & Wolters, P. S. *Dev. Brain Res.* **2**, 583–589 (1982).
32. Gryniewicz, G., Poenie, M. & Tsien, R. Y. *J. Biol. Chem.* **260**, 3440–3450 (1985).

An inositol tetrakisphosphate-containing phospholipid in activated neutrophils

Alexis E. Traynor-Kaplan, Anna L. Harris,
Barbara L. Thompson*, Palmer Taylor*
& Larry A. Sklar

Department of Immunology, Research Institute of Scripps Clinic,
La Jolla, CA 92037, USA

* Department of Pharmacology, University of California, San Diego,
La Jolla, CA 92093, USA

Inositol (1,4,5)trisphosphate ($InsP_3$)¹ and tetrakisphosphate ($InsP_4$)² have been observed in a variety of cell types and have been proposed to play roles in the receptor-mediated rise in intracellular Ca^{2+} (refs 2, 3). Recently, they have been shown to act synergistically in the activation of a Ca^{2+} -dependent K^+ channel in lacrimal acinar cells³. $InsP_3$ is the product of phospholipase C (PLC) action on phosphatidylinositol 4,5-bisphosphate ($PtdInsP_2$), whereas $InsP_4$ is believed to arise from phosphorylation of $InsP_3$ by a cytosolic kinase⁴. Although sought as a source for $InsP_4$, $PtdInsP_3$ has not been identified in any specific cell type². There were early reports of $InsP_4$ -containing phospholipids in crude extract from bovine brain⁵, but this finding was later withdrawn⁶. Recently, however, a membrane-bound enzyme (Type 1 PI kinase) which adds phosphate onto the 3 position of inositol phospholipids has been identified⁷ and the phosphatidylinositol-3-phosphate ($PtdIns(3)P$) product characterized. This suggests that several forms of phosphoinositides may exist and could be precursors for some of the variety of soluble inositol phosphate products which have been reported in recent years. Here we report the appearance of another novel phosphoinositide containing four phosphates, phosphatidylinositol trisphosphate ($PtdInsP_3$) which we find only in activated but not in unstimulated neutrophils from human donors.

Neutrophils are activated by a series of physiologic chemoattractants including formyl peptide and LTB_4 (refs 8, 9). Activation events include morphological polarization, actin polymerization, followed by depolymerization, a transient rise in intracellular Ca^{2+} , increased adhesiveness, release of lysosomal enzymes and superoxide (O_2^-) radicals. Recently the inadequacies of

Received 28 March; accepted 7 June 1988.

1. Jahr, C. E. & Jessell, T. M. *Adv. Pain Res. Therapy* **9**, 31–39 (1985).
2. Mayer, M. L. & Westbrook, G. L. *Prog. Neurobiol.* **28**, 197–276 (1987).

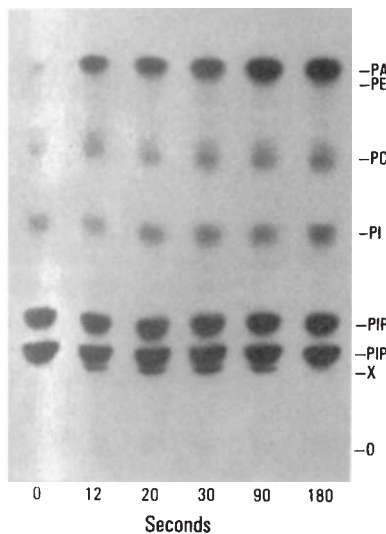


Fig. 1 Autoradiogram of TLC separation of neutrophil phospholipids showing the time course of phosphoinositide metabolism in neutrophils stimulated with 10 nM FLPEP (N-formyl-norLeu-Leu-Phe-norLeu-Tyr-Lys-fluorescein). This is typical of 18 time courses in 7 separate experiments. Abbreviations: PA, phosphatidic acid; PE, phosphatidyl ethanolamine; PC, phosphatidyl choline; PI, phosphatidyl inositol; PIP, phosphatidyl inositol 4-phosphate; PIP₂, phosphatidylinositol 4,5 bisphosphate.

Methods. Neutrophils from healthy human donors were prepared by elutriation as described²¹. This method resulted in neutrophils of >95% purity and 95% viability. Elutriated cells were washed in a Geys Buffer without carbonate containing 1.53 mM Ca²⁺, pelleted and resuspended in Buffer A (30 mM HEPES, 110 mM NaCl, 10 mM KCl, 1 mM MgCl₂, 10 mM glucose and 2 mg ml⁻¹ BSA, pH 7.4, to a density of 5 × 10⁷ cells ml⁻¹). ³²P-orthophosphate (HCl-free NEN, 500 μCi ml⁻¹) was added and the cells were incubated at 37 °C for 60 min. Cells were washed three times in Buffer A without BSA and were resuspended to 2 × 10⁷ ml⁻¹ in the same buffer with Ca²⁺ (1.53 mM). The cells were equilibrated for 5 min at 37 °C before stimulation with FLPEP (10 nM). Aliquots of 10⁷ neutrophils were added to 3 ml of chloroform/methanol (1/2, v/v) containing the antioxidant BHT (0.63 mg ml⁻¹) and phosphoinositides (Sigma, 10 μg ml⁻¹). Phospholipids were extracted essentially according to Schacht²². The lipid extract was taken to dryness under N₂ and resuspended in 100 μl chloroform/methanol (2/1, v/v) for spotting on oxalate-impregnated TLC plates (Merck, Silica Gel 60). TLC plates were developed in chloroform/acetone/methanol/acetic acid/water, (80/30/26/24/14, v/v/v/v/v).

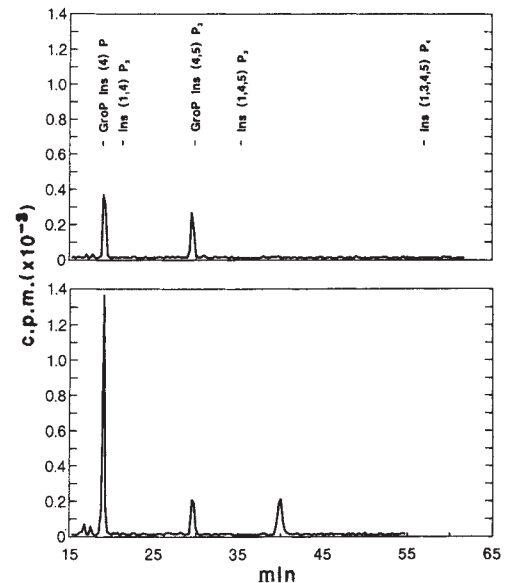


Fig. 2 HPLC eluate profiles of deacylated phospholipids coeluting with 'X' as shown in Fig. 1. The novel phosphoinositide, X, was eluted from TLC plates and deacylated by methanolysis. The deacylation products from ³²P-labelled lipids from control cells are compared with those stimulated with FLPEP for 45 s. Samples were collected every 0.25 min and counted in a Beckman LS3801. The ³²P-labelled eluate peaks were compared to retention times for ATP, ADP, and AMP and ³H-labelled internal standards (NEN): D-[inositol-2-³H(N)] 1-phosphate, D-[inositol-2-³H(N)] 1,4 bisphosphate (Ins 1,4 P₂), D-[inositol-2-³H(N)] 1,4,5 trisphosphate (Ins 1,4,5 P₃), D-[inositol-2-³H(N)] 1,3,4,5 tetrakisphosphate (Ins 1, 3, 4, 5 P₄), D-[inositol-2-³H(N)] 4,5 bisphosphate, (GroPIns(4,5)P₂).

Methods. After drying the eluate from the TLC plates under N₂, it was resuspended in 1 ml of chloroform/methanol (2/1, v/v). Na methoxide (1 ml; generated by adding 250 mg of Na metal scraped clean of oxide to 50 ml of methanol) was added to the mixture and incubated for 30 min at room temperature. Methanolysis was stopped by the addition of 1 ml of water, creating two phases. The upper phase containing the aqueous products of methanolysis was removed, the lower phase washed with 2 ml of Folch's upper layer; the upper phases were combined. Methanol was evaporated from the combined upper phases under N₂ and the remainder neutralized with formic acid. Mannitol (1 mg) was added to reduce sticking to glass surfaces. Samples in ~1.5 ml of water were loaded onto a Partisil 10 SAX HPLC anion-exchange column (4 mm × 25 cm) as described²³. This chromatogram is representative of four experiments; ³H-myoinositol-labelled lipid products show a similar profile.

previously-proposed signalling pathways in mediating these events have come into focus for several reasons. First, inhibitor studies have called into question the identities of the mediators themselves. For example, direct stimulation of protein kinase C (PKC) with phorbol esters produces many of the responses characteristic of neutrophil activation including O₂⁻ production. But PKC inhibitors are poor inhibitors of receptor-stimulated neutrophil responses¹⁰ although they inhibit phorbol ester-stimulated responses. Second, there is a lack of temporal correspondence between responses and the putative mediators. For example, the gradual, prolonged elevation of diacylglycerol (DAG) which occurs in response to formyl peptide does not correspond to the kinetics of most of the cellular events it is proposed to trigger¹¹. Third, formyl peptide stimulated O₂⁻ production was shown to be independent of Ca²⁺ elevation in intact cells¹², and Ca²⁺ and PKC activation in electroporated neutrophils, indicating that the respiratory burst is mediated through a pathway as yet undescribed¹³. Finally, no mediators for cytoskeletal activation have been unambiguously identified. Here, we show the appearance of a novel phosphoinositide,

likely to be PtdInsP₃, which may participate in mediating cell activation.

Neutrophils, stimulated with a hexapeptide agonist for the formyl peptide receptor, FLPEP (10 nM)¹⁴, transiently generate a novel phospholipid over a three-minute time course. Figure 1 is an autoradiogram of a TLC plate where ³²P-labelled phospholipids from neutrophils treated with FLPEP for 0–3 min have been separated. Note the appearance of ³²P-labelled material (X) with an R_f slightly slower than PtdInsP₂ appearing at 12 s and disappearing by 180 s. When phospholipid-containing bands were quantitated by scintillation counting, characteristic changes were observed in phosphatidic acid (PA) and PtdInsP₂. PA increased over the period of the time course in which there was a transient decrease in PtdInsP₂. Radioactivity in band X was found to increase 5–8 fold over control values at peak production (45–60 s).

The identity of X was investigated using high performance liquid chromatography (HPLC) of deacylated phospholipids. We have identified an inositol phosphate coeluting with Ins (1,

3, 4, 5) P_4 in the hydrolysates of phospholipids deacylated by two separate methods from activated but not from unstimulated neutrophils.

One method of deacylation uses sodium methoxide. On four separate occasions eluates from TLC scrapings of 'X' from phospholipid extracts from FLPEP stimulated and control cells (unstimulated) were subjected to methanolysis and the deacylated products separated by HPLC. Methanolysis is expected to cleave fatty acid chains from the lipid leaving the polar headgroup moiety attached to the glycerol. Three distinct peaks could be identified with 3H -myoinositol or $^{32}PO_4$ labelling (Fig. 2). Of the total ^{32}P radioactivity, ~51–68% eluted at 18 min with glycerol-3-phosphoinositol-4 phosphate (GroPInsP), the aqueous methanolysis product of PtdInsP. Another 13–16% eluted with glycerol-3-phosphoinositol(4,5)bispophosphate (GroPIns(4,5) P_2) at 29 min. The third peak, accounting for 8–21% of the total counts in X appeared at 39 min after the elution of the known inositol trisphosphate isomers. This peak was collected and the glycerol was cleaved from the inositol phosphates¹⁵. The product was subjected to HPLC (Fig. 3). At this time 98% of the radioactivity eluted with the Ins(1,3,4,5) P_4 standard indicating that the 39-min peak in Fig. 2 was glycerolphosphoinositol trisphosphate (GroPIns P_3).

Methylamine was also used to deacylate phospholipids. Neutrophils labelled with both $^{32}PO_4$ and 3H -myoinositol were subjected to phospholipid extraction procedures. If the source of the Ins P_4 is PtdIns P_3 we should be able to measure it directly in phospholipid extracts from stimulated neutrophils, bypassing the TLC step. This would avoid the problem of incomplete elution of the lipid from the silica gel. The fatty acids were removed by the deacylation procedure of Clarke and Dawson¹⁶ using methylamine instead of Na^+ methoxide. A portion of the deacylation product was saved for HPLC and the remainder was treated with 1 mM Na periodate for 75 min to hydrolyse the glycerol¹⁵. The sample taken before glycerol hydrolysis but after deacylation contained GroPIns4P and GroPIns(4,5) P_2 and the FLPEP-stimulated sample contained a peak at 39 min, (where GroPIns P_3 elutes) but there was no radioactivity eluting with Ins(1,3,4,5) P_4 . Therefore, soluble Ins P_4 was not carried along during the extraction procedure. After the glycerol was hydrolysed, ~89% of the GroPIns4P and GroPIns(4,5) P_2 was converted to Ins(1,4) P_2 and Ins(1,4,5) P_3 respectively. Of the counts in the 39 minute peak, 80% were converted to counts coeluting with Ins(1,3,4,5) P_4 (57–60 min).

In a separate experiment neutrophils were double-labelled with $^{32}PO_4$ and 3H -myoinositol. Phospholipids were extracted and deacylated as described in the figure legends. Aliquots were subjected to HPLC to ascertain that no soluble inositol phosphates were carried along during the extraction procedure. The remainder were subjected to hydrolysis of the glycerol. The products were then subjected to HPLC (Fig. 4). The ^{32}P and 3H -myoinositol labelled material coelute exactly demonstrating that they are both incorporated into the same material.

Although it seems clear that part of the radioactivity in the band below PtdIns P_2 comes from the Ins P_4 -containing compound in stimulated neutrophils, how do we account for the presence of GroPInsP and GroPIns P_2 containing material in this region? Because these substances are also in the control samples where no Ins P_4 -containing material was seen, it is unlikely that they are breakdown products of Ins P_4 or GroPIns P_3 . Furthermore, we have recently found that lyso-PtdInsP comigrates with the Ins P_4 -containing phospholipid on TLC, explaining the presence of Ins P_2 . The Ins P_3 probably results from incomplete resolution from the PtdIns P_2 , which migrates very close to X. As lysoPtdIns P_2 migrates with a slower R_f than X the Ins P_4 -containing phospholipid is probably not lysoPtdIns P_3 , which would be expected to migrate with or below lysoPtdIns P_2 . The TLC and HPLC data are consistent with the hypothesis that the Ins P_4 -containing molecule is phosphatidylinositol(3,4,5) phosphate (PtdIns(3,4,5) P_3), as the inositol phos-

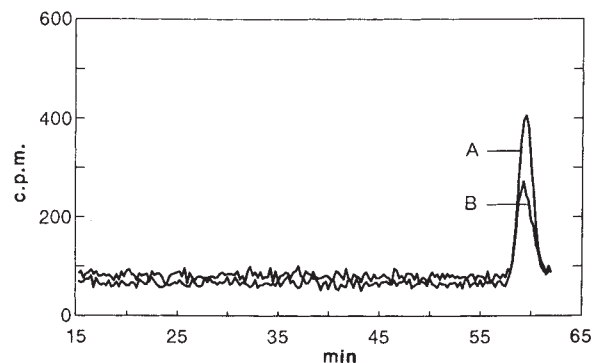


Fig. 3 HPLC of hydrolysed eluate from the 39-min peak as shown in Fig. 2. Fractions eluting between 38.5 and 39.25 min were collected from the HPLC and neutralized with cyclohexylamine. The glycerol was hydrolysed as described¹⁵. The product of this procedure was run on HPLC as in Fig. 2. Samples were counted in a Searle 6890 Liquid Scintillation Counter. A, sample from $^{32}PO_4$ -labelled neutrophils stimulated for 45 s with FLPEP (10 nM). B, 3H Ins (1,3,4,5) tetrakisphosphate standard (NEN).

phate recovered from its hydrolysis coelutes with Ins(1,3,4,5) P_4 . As there are two other known isomers of Ins P_4 which elute very close to Ins(1,3,4,5) P_4 (refs 17, 18) we cannot be certain of its precise substitution pattern until we complete further analysis.

Several observations indicate the occurrence of PtdIns P_3 may be a physiologically relevant finding. First, the appearance of this lipid only after stimulation suggests that receptor activation in neutrophils may be linked to stimulation of a PtdIns P_2 kinase. It does not appear in neutrophils pretreated with pertussis toxin, indicating that its formation is linked to a G protein (manuscript in preparation). PtdIns P_3 could arise from the action of a kinase on PtdIns(4,5) P_2 . The PtdIns(3) kinase⁷ has been found to act on commercially available PtdPIns(4,5) P_2 , creating PtdIns P_3 (L. Cantley, personal communication). Second, the respiratory burst reportedly can be stimulated through the formyl peptide receptor without a rise in intracellular Ca^{2+} or PKC activation, two signals previously thought to be involved in O_2^- production¹³. The appearance of radioactivity in the band containing PtdIns P_3 has the same time course as the rate of O_2^- production unlike other putative intracellular mediators in neutrophils. Third, mediators for cytoskeletal changes remain unidentified in neutrophils. As a highly charged phospholipid, PtdIns P_3 could participate in anchoring membrane proteins and the con-

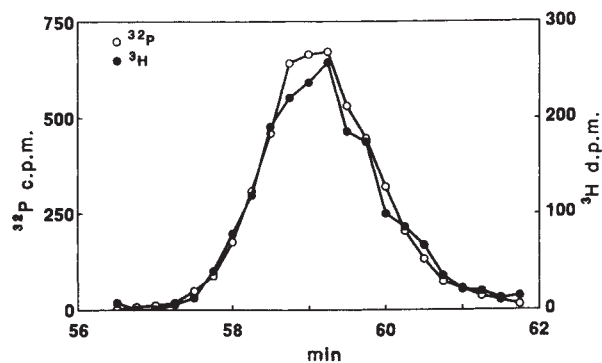


Fig. 4 $^{32}PO_4$ and 3H -myoinositol label the same entity. Neutrophils were double-labelled with $^{32}PO_4$ and 3H -myoinositol using the incubation conditions described in Fig. 1 except that cells were incubated with 3H -myoinositol for 3 h before addition of $^{32}PO_4$, and incubated another hour. Phospholipids were extracted, deacylated and glycerol hydrolysed as in Fig. 3.

trol of membrane cytoskeletal interactions¹⁹. Finally, early increases in InsP₄ and InsP₅ before the rise in Ins(1,4,5)P₃ in TRH-stimulated GH₃ cells²⁰ suggests there are other routes for the InsP₄ production than through the cytosolic InsP₃ kinase.

Received 3 May; accepted 6 June 1988.

1. Streb, H., Irvine, R. F., Berridge, M. J. & Schultz, I. *Nature* **306**, 67–69 (1983).
2. Irvine, R. F. & Moor, R. M. *Biochem. J.* **240**, 917–920 (1986).
3. Morris, A. P., Gallacher, D. V., Irvine, R. F. & Petersen, O. H. *Nature* **330**, 653–655 (1987).
4. Irvine, R. F., Letcher, A. J., Heslop, J. P. & Berridge, M. J. *Nature* **320**, 631–634 (1986).
5. Santiago-Calvo, E., Mulé, S. J. & Hokin, L. E. *Biochim. biophys. Acta* **70**, 91–93 (1963).
6. Santiago-Calvo, E., Mulé, S., Redman, C. M., Hokin, M. R. & Hokin, L. E. *Biochim. biophys. Acta* **84**, 550–562 (1964).
7. Whitman, M., Downes, C. P., Keeler, M., Keller, T. & Cantley, L. *Nature* **332**, 644–646 (1988).
8. Goldstein, I. *et al. Proc. natn. Acad. Sci. U.S.A.* **70**, 2916–2929 (1973).
9. Malmsten, C. L. *et al. Acta physiol. scand.* **11**, 449–451 (1980).
10. Wright, C. G. & Hoffman, M. D. *Biochem. biophys. Res. Comm.* **142**, 53–62 (1987).
11. Rider, L. G. & Neidel, J. E. *J. biol. Chem.* **262**, 5603–5609 (1987).
12. Lew, P. D., Wollheim, C. B., Waldvogel, F. A. & Pozzan, T. *J. cell Biol.* **99**, 1212–1220 (1984).
13. Grinstein, S. & Furuya, W. *J. biol. Chem.* **263**, 1779–1782 (1988).
14. Sklar, L. A. *et al. J. biol. Chem.* **259**, 5661–5669 (1984).
15. Hawkins, P. T., Michell, R. H. & Kirk, C. J. *Biochem. J.* **218**, 785–793 (1984).
16. Clarke, N. G. & Dawson, R. M. C. *Biochem. J.* **195**, 301–306 (1981).
17. Stephens, L. *et al. Biochem. J.* **249**, 271–282 (1988).
18. Balla, T., Guillemette, G., Baukal, A. J. & Catt, K. J. *J. biol. Chem.* **262**, 9952–9955 (1987).
19. Janney, P. A., Iida, K., Yin, H. L. & Stossel, T. P. *J. biol. Chem.* **262**, 12228–12236 (1987).
20. Heslop, J. P., Irvine, R. F., Tashjian, A. H. & Berridge, M. J. *exp. Biol.* **119**, 395–401 (1985).
21. Tolley, J. O., Omann, G. M. & Jesaitis, A. J. *J. Leuk. Biol.* **42**, 43–50 (1987).
22. Schacht, J. *J. Lipid Res.* **19**, 1063–1067 (1978).
23. Ambler, S. K., Thompson, B., Solski, P. A., Brown, J. H. & Taylor, P. *Molec. Pharmac.* **32**, 376–383 (1987).

New features of microtubule behaviour observed *in vivo*

Eric Schulze & Marc Kirschner

Department of Biochemistry and Biophysics,
University of California, San Francisco, California 94143-0448, USA

The microtubule cytoskeleton is thought to be intimately involved in generating and maintaining cell polarity^{1,2} and can generate many different morphological structures from a few structural elements³. The mechanism by which these structures are generated has been partially elucidated from studies of microtubule polymerization both *in vitro*^{4–6} and *in vivo*^{7,8}. Microtubules *in vitro* exist in growing (polymerizing) and shrinking (depolymerizing) populations that interconvert infrequently. This behaviour, termed dynamic instability⁴, permits microtubules in the cell rapidly to explore different arrangements and allows selective stabilization of specific morphologies^{9,10}. To investigate the regulation of these processes, we have implemented techniques for direct observation of fluorescently labelled microtubules¹¹ and developed them to observe the dynamic behaviour of individual microtubules in single living cells. Sammak and Borisy¹² recently used this technique to show that the dynamics of microtubules in fibroblasts is explained by dynamic instability. Although we also conclude here that dynamic instability explains much of microtubule behaviour *in vivo*, we find significant deviations from the properties of tubulin *in vitro*. These results suggest that local cytoplasmic factors strongly influence microtubule dynamics; such control has important implications for cellular morphogenesis.

We used fluorescent labelling of proteins coupled with low-light-level video microscopy and computer image processing to see individual microtubules in living cells for extended periods of time. Figure 1 shows several frames of individual microtubules at the periphery of a living fibroblast cell. To obtain satisfactory continuous recordings, we injected African green monkey kidney (BSC1) or Potorous tridactylis kidney (Ptk1) cells with rhodamine-labelled tubulin, and incubated for 1–2 h at 33 °C to allow the cell to recover from any perturbations induced by the injection and to allow uniform incorporation of the rhodamine-

labelled tubulin into the vast majority of microtubules. These rhodamine-tubulin preparations have been shown previously to function normally in *in vitro* assembly experiments, and to participate in both mitotic and interphase microtubule arrays in *Drosophila* embryos for many cell cycles¹¹.

We chose a small area at the periphery of an injected cell, where microtubules are sparse, for study. This area was viewed using epifluorescence at very low light levels, applied intermittently, through an aperture that limits illumination to a small region of the cell (see Fig. 1). Higher levels of illumination or low light levels applied continuously lead to marked changes in microtubule dynamics and cell behaviour. This is manifest by diminished microtubule dynamics, retraction of cell processes, and, over a long period of time, cell death. At low levels of illumination at 33 °C (cells are more resistant to photodamage at 33 than at 37 °C) we see no differences in microtubule growth rates or shrinkage rates from the beginning to the end of a sequence, often lasting 1 h; thus our observations are probably not significantly affected by photodynamic damage. Because the image from the intensified silicon-intensified target camera is very noisy at these low light levels, we use a digital image processor to average 10 frames (~1/3 s) before storing on optical disk.

Figure 1(a–f) shows six frames of a time-lapse recording of microtubules observed at the edge of a Ptk1 cell. These frames, taken 40 s apart, demonstrate the direct observation of an individual microtubule and its behaviour over time. The arrows point to the end of a microtubule and to a fixed reference point; the microtubule grew 4.7 µm in the 3 min 14 s of observation included in this sequence. The average growth rate for this microtubule was 1.5 µm per min and the range of instantaneous growth rates was 1.3–1.9. Other microtubules in this sequence showed both growing and shrinking transitions. Both Ptk1 and BSC1 cells show a broad range of growth rates but have similar mean rates (Fig. 1g, h). The error in measurement corresponds to a standard deviation of only 7%. The measurements were made in very flat regions of cells where the non-growing segments do not enter and leave the plane of focus and hence we do not anticipate that blurring because of focus problems contributes large errors to the measured growth rates.

To investigate the cause of the wide range of growth rates, we observed microtubules at shorter time intervals to see if there were periodic transitions between stopping or shrinking. We found no evidence for short transitions; the instantaneous rates of growth were the same at 5-s as at 20-s intervals. To address this question of whether variability in growth rates is caused by variation within a single microtubule over time or between different microtubules at a given time, we plotted consecutive microtubule growth rates for individual microtubules against each other (Fig. 1i). This scatter plot of consecutive growth rates reveals only a very small correlation between a microtubule's new and its previous growth rates suggesting that microtubules show very little persistence in their rate of growth.

Transitions between the phases of growth, stability and shrinkage are infrequent (typically occurring less than once per 5 µm of growth, or 8,000 monomer units), but they seem to occur more frequently than anticipated. The exact frequency of transitions is hard to estimate, as microtubules usually transit out of the picture during the depolymerization phase. Some examples of different types of individual microtubule histories (length versus time plots) are given in Fig. 2. These examples reflect the range of behaviours, but not the frequencies with which each type occurs. In addition to alternating growing and shrinking phases, there are also transitions to stable or metastable states.

Previous studies of microtubule dynamics *in vivo* were unable to measure directly the rate of microtubule shrinkage, and relied instead on estimating the bulk shrinkage rate indirectly as that rate needed to balance the fraction of growing microtubules and their average rate of growth. Direct observation of shrinkage

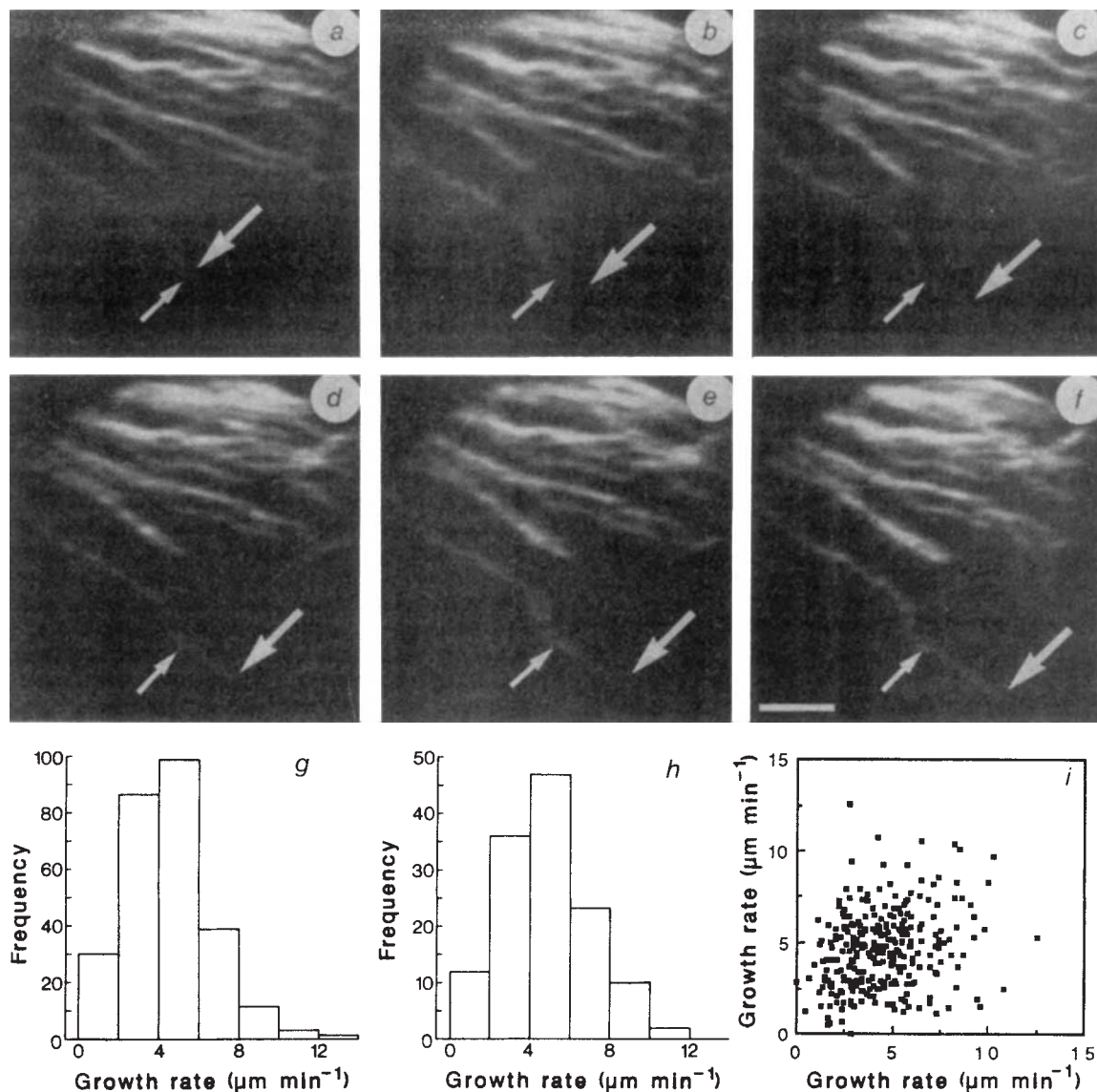


Fig. 1 *a-f*, Six successive frames of a time-lapse recording of individual rhodamine-labelled microtubules in a live Ptk1 cell. This cell was microinjected with 6 mg ml^{-1} rhodamine-labelled tubulin¹¹ 1 h before visualization using low-light-level video microscopy. Time interval between frames, 39 s. Large arrow, microtubule end; small arrow, position of the microtubule end at time 0. The length of this microtubule increases $4.7 \text{ } \mu\text{m}$ in the 3 min 14 s covered in this sequence. Average rate of growth of this microtubule is therefore $1.5 \text{ } \mu\text{m min}^{-1}$. *g*, Histogram of 269 growth rates measured in 5 separate Ptk1 cells incubated at 33°C . Average growth rate: $4.5 \text{ } \mu\text{m min}^{-1}$; s.d. 2.1; range, $0.5\text{--}12.5 \text{ } \mu\text{m min}^{-1}$; s.e. 0.13. *h*, Histogram of 131 growth rates measured in 4 BSC1 cells incubated at 33°C . Average growth rate: $5.0 \text{ } \mu\text{m min}^{-1}$; s.d. 2.4; range, $0.4\text{--}12.0 \text{ } \mu\text{m min}^{-1}$; s.e. 0.21. *i*, Measured consecutive microtubule growth rates. Each point represents a growth rate measured at one time interval plotted against the growth rate for the same microtubule measured during the previous time interval. The slope of the best-fit line is 0.24 with a correlation coefficient of 0.25. The 95% confidence range, 0.24 ± 0.20 . Hence the correlation between growth rates is only very slightly greater than 0–95% confidence. (Bar, $5 \text{ } \mu\text{m}$.)

Methods. To image microtubules *in vivo*, Ptk1 and BSC1 cells grown on glass coverslips are fitted into a 33°C temperature-controlled chamber mounted on a Zeiss ICM 405 inverted microscope and microinjected⁴ with rhodamine-labelled tubulin¹¹. After $\sim 1 \text{ h}$ of *in situ* incubation at 33°C , a small area near the periphery of an injected cell is then identified for further study. A shuttered halogen lamp (Zeiss 12 V, 100 W) supplies light of variable intensity (typically set to provide 0.008 mW cm^{-2}) which passes through the microscope's adjustable epifluorescence aperture to provide rhodamine epifluorescence illumination (using the standard Zeiss rhodamine epifluorescence filter set) to only the area under observation (usually $10 \text{ } \mu\text{m}$ diameter). Time-lapse video sequences are obtained using an Interactive Video Systems (IVS) Image-1 digital image processor to record 1/3-s averaged images from a COHU (San Diego) intensified silicon intensified target camera (ISIT, model 5162-2000-0000) every 10 s to a panasonic model TQ 2025F high-resolution monochrome optical memory disk recorder. A cell so observed for 1 h will thus receive only the equivalent of 2 min of illumination over $<10\%$ of its area. The sequences so obtained were then played back in loops under computer (IBM, PC) control while growth and shrinkage rate measurements are made using a GITCO digitizing pad. The information is then stored in a computer file for subsequent statistical analysis.

rates reveals a bimodal distribution in both BSC1 and Ptk1 cells. We surmise that these different rate classes are related to the previous history of the microtubule. To test this idea, we divided the shrinkage data into those microtubules that had not grown or shrunk for more than 40 s (stable or metastable microtubules) and those that had been dynamic. When plotted on separate histograms (Fig. 3) we find that previously stable microtubules in Ptk1 cells shrink half as fast as microtubules that had transitioned directly from the growing phase. Similar data were obtained for

BSC1 cells. This difference in microtubule shrinkage rate probably results from a modification along the whole length of the microtubule, such as the binding of proteins or post-translational modification of tubulin.

In addition to obtaining quantitative data on microtubule growth and shrinkage rates, the direct observation of microtubules in living cells produces interesting observations about microtubule behaviour. Some microtubules show no growth or shrinkage during the entire observation (30–60 min) and the

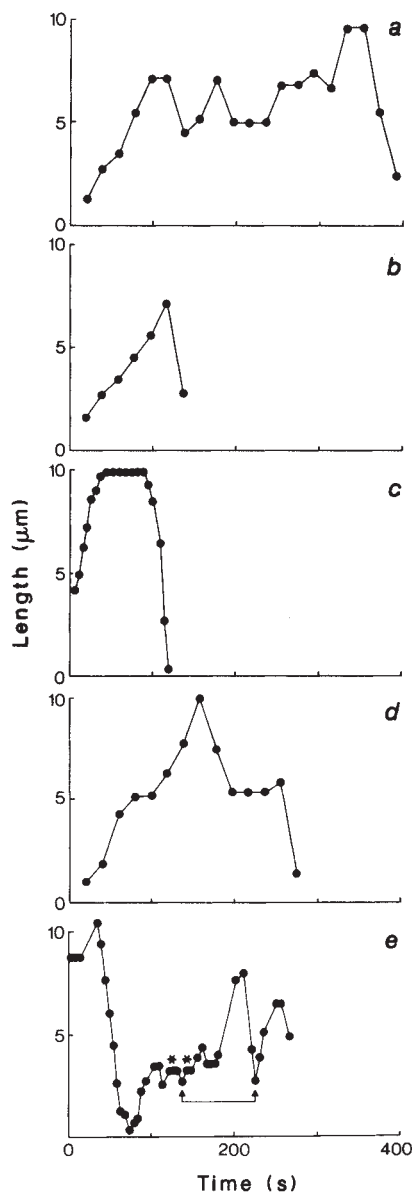


Fig. 2 Length versus time graphs for five individual microtubules in Ptk cells. The cells were incubated at 33 °C during visualization. Inspection of the microtubule histories reveals a variety of behaviours. The microtubule in *e* makes numerous transitions from growth to shrinkage. Paired arrows and asterisks in *e* point to two sites along the polymer where a transition occurs twice.

frequency of such microtubules seems to be different in different regions of a cell. These microtubules presumably correspond to the stable population observed previously¹³. Although most microtubules appear to grow randomly in the cytoplasm, there are clear examples where a new microtubule accurately followed the path of a previous microtubule. Figure 4 illustrates one such example. In this sequence a microtubule that recently grew out to the cell edge (Fig. 4*a*) depolymerizes (Fig. 4*b*). A second microtubule then grows out and follows the same curved path as the first. The complete video sequence strongly reinforces the impression that the two microtubules follow exactly the same path (Fig. 4*d*).

The results from our experiments confirm previous results with biotin-tubulin injections in terms of the mean rate of growth and the dispersion in growth rates, and suggest that the biotin tubulin technique does not suffer from problems attributable to the trauma of injection. Dynamic instability *in vitro* has also

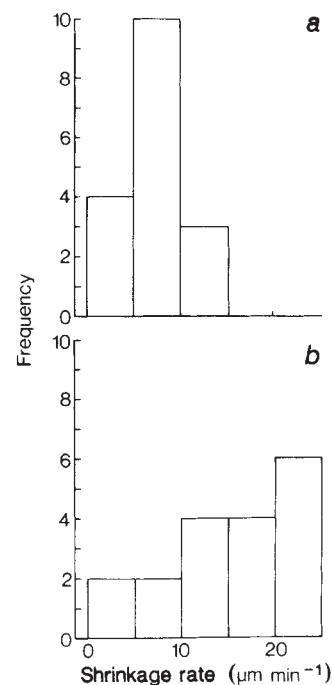


Fig. 3 Histograms of microtubule depolymerization rates in Ptk cells incubated at 33 °C during visualization. *a*, Histogram of depolymerization rates for microtubules which were stable for 40 s or more before depolymerizing (mean rate of depolymerization, $7.4 \pm \text{s.d. } 2.7$, s.e. $0.64 \mu\text{m min}^{-1}$). *b*, Histogram of depolymerization rates for microtubules which transited directly from growth to shrinkage (mean rate of depolymerization is $14.0 \pm \text{s.d. } 6.1$, s.e., $1.4 \mu\text{m min}^{-1}$).

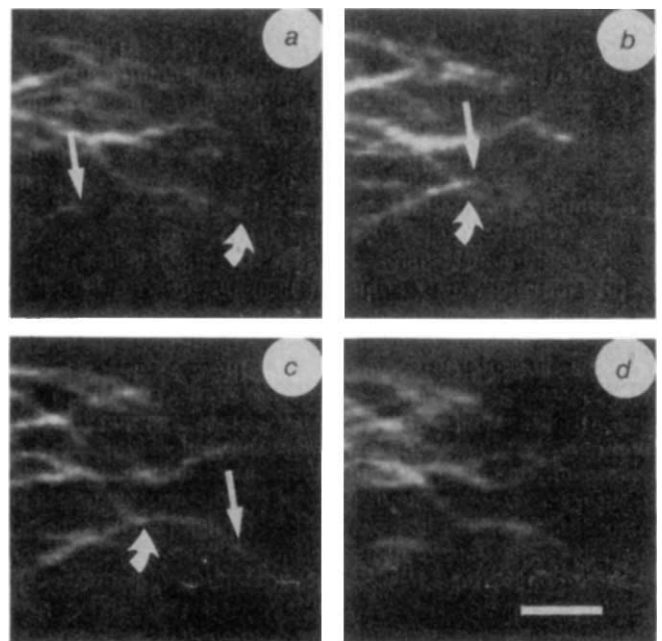


Fig. 4 Illustration of microtubule tracking in a Ptk1 cell (33 °C). The microtubule indicated by the curved arrow in *a* depolymerizes to the point indicated by curved arrow in *b*; meanwhile a second microtubule, indicated by the straight arrow in *a-c* grows out following the same path as the first microtubule (*c*). *d*, Photographic superposition of *a* and *c* demonstrating the identity of the two tracks. (Bar, 4 μm .)

been recently confirmed¹⁴. Our present studies, however, demonstrate new features that were not expected based on *in vitro* experiments and that were not predicted from theoretical studies of pure tubulin. Growth rates are variable *in vivo* and this variability far exceeds the statistical variability expected if tubulin adds as single monomer units. There appears to be no persistence to the growth rates and at every instant a growing microtubule seems to be affected by a different immediate environment. Shrinkage rates, by contrast, although variable, show some persistence. Microtubules that had been stable before initiating depolymerization depolymerize more slowly, while those which transited directly from a growth phase to a shrinkage phase shrink roughly twice as fast. This may be because of microtubule-associated protein binding or post-translational modification^{15,16}, which may in turn reflect microtubule age¹⁷.

We find that the transition probability from shrinking to growth is not zero, so that polymers could reverse direction before complete depolymerization. The overall contribution of this reversal to the microtubule flux is difficult to estimate, as many shrinking microtubules depolymerize so extensively that they are lost from view. We notice that some microtubules seem to reverse direction at specific locations (paired arrows and asterisks in Fig. 2). Occasionally we notice three reversals at the same site, suggesting that there is a stabilizing structure at that point. Microtubules also track along the same path occupied just previously by other microtubules, suggesting that physical barriers or materials which promote assembly, such as microtubule-associated proteins, remain localized after the depoly-

merization of a microtubule; localized structures such as these could have a significant influence on microtubule growth and distribution. These new features of microtubule behaviour, produced by modulation of microtubule dynamics and control of microtubule stability by local cytoplasmic factors, could be important in cellular morphogenesis.

We thank Tim Mitchison and Andrew Murray for discussions and Tim Mitchison for developing and refining the rhodamine-tubulin fluorescent probe. We also thank Jon Minden for technical assistance, and Irma Delson for help with preparation of the manuscript. Supported by grants from the NIGMS.

Received 15 April; accepted 9 June 1988.

1. Goldman, R., Pollard, T. D. & Rosenbaum, J. *Cell Motility* (Cold Spring Harbor, New York, 1976).
2. Gotlieb, A. I., May, L. M., Subrahmanyam, L. & Kalnins, V. I. *J. cell Biol.* **91**, 539–594 (1981).
3. Inoue, S. in *40th Symp. Soc. Dev. Biol.* (eds Subtelny, S. & Green, P. B.) 35–76 (Liss, New York, 1976).
4. Mitchison, T. J. & Kirschner, M. W. *Nature* **312**, 232–237 (1984).
5. Mitchison, T. J. & Kirschner, M. W. *Nature* **312**, 237–241 (1984).
6. Kristofferson, D., Mitchison, T. J. & Kirschner, M. W. *J. cell Biol.* **102**, 1007–1019 (1981).
7. Schulze, E. & Kirschner, M. *J. Cell Biol.* **102**, 1020–1031 (1986).
8. Saxton, W. M. *et al. J. cell Biol.* **99**, 2175–2186 (1984).
9. Kirschner, M. & Schulze, E. *J. cell Sci. Suppl.* **5**, 293–310 (1986).
10. Kirschner, M. & Mitchison, T. J. *Cell* **45**, 329–342 (1986).
11. Kellog, D. R., Mitchison, T. J. & Alberts, B. M. *Development* (in the press).
12. Sammak, P. J. & Borisy, G. G. *Nature* **332**, 724–726 (1988).
13. Schulze, E. & Kirschner, M. *J. cell Biol.* **104**, 277–288 (1987).
14. Salmon, E. D. *et al. 45th Ann. Proc. Electron Microscopy Soc. Am.* (ed. Bailey, G. W.) 636–637 (San Francisco, San Francisco, 1987).
15. Thompson, W. C., Asai, D. J. & Carney, D. H. *J. cell Biol.* **98**, 1017–1025 (1984).
16. Gunderson, G. G., Kalnmoski, M. H. & Bulinski, J. C. *Cell* **38**, 779–789 (1984).
17. Schulze, E., Asai, D. J., Bulinski, J. C. & Kirschner, M. *J. cell Biol.* **105**, 2167–2177 (1987).

Structure and elasticity of microtubule-associated protein tau

B. Lichtenberg, E.-M. Mandelkow, T. Hagestedt & E. Mandelkow*

Max-Planck-Unit for Structural Molecular Biology, c/o DESY, Notkestrass 85, D-2000 Hamburg 52, FRG

Tau is one of the diverse group of microtubule-associated proteins that bind to microtubules and may thereby influence their structure and function. It occurs in the mammalian brain¹, mainly in axons, and is a component of the neurofibrillary tangles of Alzheimer's disease^{2,3}. Tau was recently sequenced⁴, but there remains a shortage of structural data on the protein. We have now prepared paracrystals of tau suitable for electron microscopy and image processing. They show distinct transverse banding and polarity, indicating that the protein subunits are aligned with the same orientations. In contrast to other paracrystals, those of tau protein can stretch or contract continuously by more than three-fold; the axial repeats range from 22 to 68 nm. After scaling to a common period, the density distributions are closely superimposable. This suggests that tau is an elastic molecule.

Microtubules are hollow cylinders made of tubulin and various microtubule-associated proteins (MAPs), including tau protein. Tau is actually a mixture of four to five related polypeptides, one of which has been sequenced⁴. It is heat stable and soluble in perchloric acid^{5,6} and these properties form the basis for the isolation procedure (see Fig. 1).

Paracrystals were grown by dialysis on an electron microscope grid (Fig. 1). They are straight or slightly curved, but highly coiled specimens are also often observed (Fig. 1d), indicating that the structures can accommodate internal stress. The diameters are typically 60–120 nm while the lengths vary up to 30 μ m, but are generally in the range of 1–3 μ m. The paracrystals

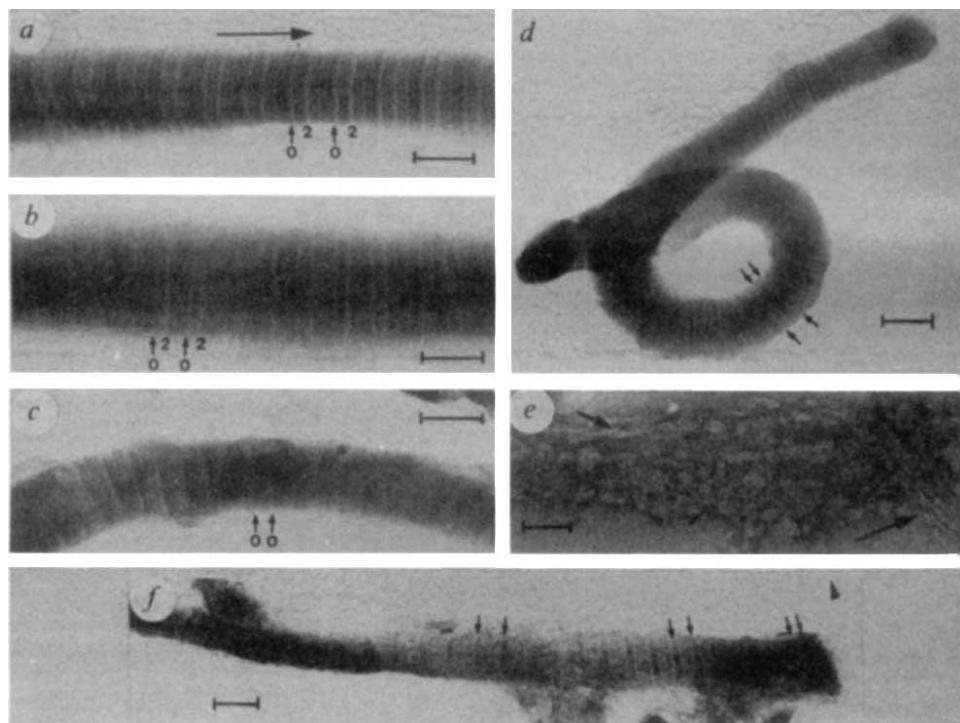
show patterns of light and dark transverse bands. This is typical of structures where elongated molecules are aligned parallel to one another along the axis, the gaps and overlaps leading to the variations in stain distribution—well-known examples are collagen⁷ or tropomyosin⁸. The crystals with the largest repeats (>60 nm) show the most detailed banding pattern, with up to seven transverse striations per repeat, separated by darker regions. The most prominent bands (termed 0 and 2, Fig. 1a, b) are separated by roughly one third of the repeat. Between them are two faint striations (bands 1a and 1b), and three more bands (3, 4, 5) occur between band 2 and 6 (equivalent to band 0). The polarity of the pattern is most easily seen from the dark 'gap' preceding band 0 in Fig. 1a. This pattern will be referred to as type I. The polarity means that the molecules must all be arranged in the same direction. In this regard the paracrystals are similar to those of collagen, but different from tropomyosin where the subunits are packed with antiparallel orientations⁸.

As the repeat decreases, the banding pattern gradually changes. This can be seen by comparing different particles (such as Fig. 1a, b), but it also occurs within a given particle when the repeat decreases, particularly near bends or towards the ends (Fig. 1d, f). The contracted pattern (type II, Fig. 1b, d) usually shows less polarity; it has two approximate twofold axes centred on bands 1 and 4. In many particles the faint striations become invisible, leaving only the dark region between bands 0 and 2 and the lighter region between 2 and 0 (type III, Fig. 1c, f), again with little indication of polarity. This pattern is often adjacent to type II patterns and appears to result from disordering. As in the type I pattern, the separation between bands 0 and 2 is roughly one third of the repeat. The rather symmetrical patterns of type II and III could, in principle, arise from an antiparallel packing of molecules. But many structures contain gradual transitions between patterns of type I and II or III, often with a corresponding change in spacing. This means that even the less polar patterns arise from molecules packed with the same orientation.

A striking feature is the variability of repeats (between 22 and 68 nm, Figs 1f, 2a) not only in different particles in the same field, but also within a given particle. Moreover, the repeat is

* To whom correspondence should be addressed.

Fig. 1 Images of tau paracrystals. The bar in Fig. 1e is 50 nm, all others are 100 nm. *a*, Type I paracrystal, periodicity 63 nm. The two most prominent bands (0 and 2) are marked. There are two faint striations between bands 0 and 2 (1a and 1b), and three striations between band 2 and the next band 0 (3, 4, 5). The dark gap to the left of band 0 allows the polarity to be observed by eye (long arrow). *b*, Type II paracrystal, periodicity 53 nm. Bands 1a and 1b are no longer resolved, and the gap preceding band 0 has largely disappeared. *c*, The centre part is a type III pattern, whereas the outer parts approach the appearance of type II. The periodicity varies around 37 nm. *d*, Coiled paracrystal with a type II banding pattern. The periodicity varies from 22 nm at the inner edge to 45 nm at the outer edge (reminiscent of an accordion). *e*, Disintegrating paracrystal. The arrows point to 'protofibrils' fraying off from the paracrystal. The arrow on the top left points to an axial periodicity of ~4–4.5 nm, that in the centre to an axial periodicity of ~8–9 nm, and that on the right shows protofibrils of apparent width 3–6 nm. *f*, Paracrystal ending roughly at the position of band 0, as determined from other parts further inside. The periodicity decreases from 56 nm inside down to 30 nm towards the end, and the particle becomes wider.



Methods. Microtubule protein from porcine brain was prepared by a modified temperature cycle method¹³ and then boiled for 15 min (ref. 5) before a clearing spin. The supernatant was then applied to a Pharmacia FPLC Mono-S column and eluted with a NaCl gradient. The protein was precipitated with 2.5% perchloric acid to remove the non-tau MAPs⁶. The supernatant containing tau was precipitated in 45% ammonium sulphate and the pellet was redissolved in 20 mM PIPES pH 6.9 with 1 mM each of EGTA, MgSO₄, dithiothreitol, and with 50 mM NaCl. SDS-PAGE showed the usual 4–5 isotypes^{1,6}. Paracrystals were grown by grid dialysis¹⁴ against 10–50 mM Na acetate buffer, pH 5.4–6.5, 20–50 mM MgCl₂, using copper grids coated with collodion/carbon. The specimens were stained with 1% uranyl acetate and examined on a Philips CM12 microscope at 35–45,000 magnification.

correlated with the thickness of the aggregate; the wider it becomes, the shorter the repeat (Fig. 2b). This is most prominent near the ends (Fig. 1f) but occurs also within the paracrystals and is reminiscent of the 'constant volume effect' in striated muscle fibres. To visualize this, imagine a piece of rubber tubing: when it is stretched, it becomes thinner; when it is compressed, it becomes thicker. The variability of repeats distinguishes tau paracrystals from others whose spacings change only within narrow limits. A related property of the paracrystals is their elasticity which allows them to bend sharply without breaking. In the example of Fig. 1d, the repeat on the outside is 45 nm, but on the inside it is only 22 nm, just half of the outside repeat. In spite of this compression the banding pattern is preserved (type II, both in the straight and coiled regions). This suggests that the molecules themselves are elastic, and the structure behaves like an accordion.

Many of the tau paracrystals contain longitudinal striations. Some are fairly broad and have lateral separations of ~18 nm (mainly between bands 3 and 5; data not shown). Others form protofibrils 3–6 nm wide which could correspond to the widths of single tau molecules or aggregates of it. They can be seen not only in the paracrystals but also in the less well ordered material (Fig. 1e). Furthermore, in selected areas along the protofibrils or at the edges of paracrystals there are axial periodicities of ~4 or 8 nm (Fig. 1e), similar to the axial periodicities of microtubules.

Because the banding pattern arises from the superposition of staggered molecules, the beginning and end of individual molecules is not defined. This problem can be addressed by observing the ends of the paracrystals. In most cases they terminate in a cloud of fuzzy material, but in a few cases it was possible

to trace the striations. For example, the particle of Fig. 1f appears to start near band 0. As the next set of molecules would start at the next band 0, the density distribution from the end to the first internal band 0 would reflect the intrinsic density of tau. We note that this distribution is similar to that found inside well ordered paracrystals, namely a dark region of about one third of the repeat between bands 0 and 2, followed by a brighter

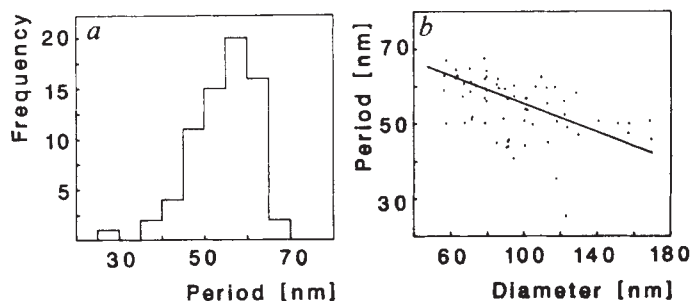


Fig. 2 Variation in axial repeat and correlation with thickness in the paracrystals. *a*, Histogram of periodicities in 5 nm bins. Most paracrystals have repeats between 50–65 nm, with a sharp upper cutoff at 68 nm. The smallest observed value is 22 nm. *b*, Correlation between axial periodicities and widths. The entries contain data from different paracrystals as well as variations in periodicity and width occurring within single paracrystals. In both cases the correlation is negative, that is wider paracrystals tend to have shorter repeats. The increase in diameter and decrease in periodicity is often most apparent at the ends (Fig. 1f) but occurs within paracrystals as well.

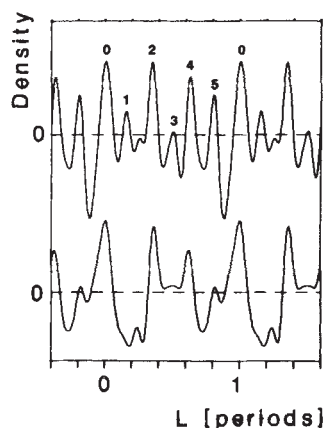


Fig. 3 One-dimensional reconstruction of the density distribution along the axis of two paracrystals. Top, type I (65 nm repeat), bottom, type II paracrystal (53 nm repeat). The bottom trace has been expanded laterally by a factor of $65/53 = 1.23$. The dashed lines show the zero levels of the reconstructions; the y-scale is arbitrary. The density peaks labelled 0–5 correspond to the transverse striations seen in Fig. 1; the highest ones are bands 0 and 2. In type I particles (top), bands 0 and 2 are separated by a trough in which up to two bands (1a and 1b, often not resolved) can be seen. The space between band 2 and the next band 0 is subdivided by three bands (3–5), with a pronounced trough between bands 5 and 0 (useful for determining the polarity by eye, as in Fig. 1a). In type II particles (bottom), the trough between bands 0 and 2 is usually deeper, and bands 1a and 1b merge into one. Bands 3 and 5 are of roughly equal height and weaker than band 4. The trough seen here between bands 4 and 5 is often less pronounced, and the trough between bands 5 and 0 of the top trace has largely disappeared. As a result, type 2 patterns tend to be less polar, with approximate twofold axes centred on bands 1 and 4. The banding patterns of all type I and II paracrystals observed thus far are superimposable in this fashion, with only moderate variations in peak heights and positions, and there is a gradual transition between them as the periodicity decreases. Type III patterns can be regarded as a less ordered variant of type II where the fine detail has been lost, that is they show a darker region of about one third of the repeat (the trough between bands 0 and 2), and a brighter region corresponding to the average of bands 2–5.

Methods. Electron micrographs were screened by optical diffraction, digitized, and computer processed (Fourier transformation, noise filtering, reconstruction). The reconstructions were displayed such that high densities correspond to protein, low densities to stain.

region between band 2 and 0 (roughly two thirds of the repeat). Thus, to a first approximation, the density from band 0 to just before the next band 0 seems to be dominated by a single set of tau molecules in register. These considerations were incorporated in a preliminary model of the structure of tau and its packing in the paracrystal (see below, Fig. 4). A more accurate determination of the molecular boundaries must await further studies (such as by antibody labelling).

Optical or computed diffraction patterns show reflections up to the 11th order, corresponding to 5–6 nm resolution. Figure 3 shows reconstructed density distributions of type I and II, scaled to the same repeat so that the 0 bands are superimposed. With this procedure it becomes apparent that the other bands are also nearly superimposed. There is some variation in peak heights and positions, but considering the change in repeat it is surprisingly small. Thus, to a good approximation the reconstructed densities can be transformed into one another simply by stretching or compressing.

The significance of this result becomes clear if we try to apply the 'standard' analysis of paracrystals. For example, the appearance of collagen fibres is dominated by the stagger between the

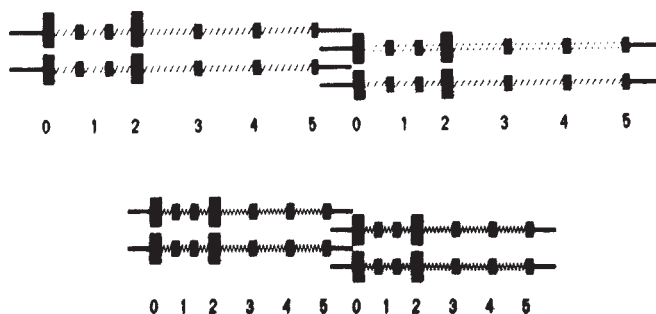


Fig. 4 Model of the structure of tau molecules and their packing in the paracrystal. The filled blocks represent high protein density (bands in the paracrystals), the springs indicate the elasticity of the molecule. The distinction between structural and elastic elements is exaggerated for clarity. The top and bottom diagrams show the molecules in an extended and contracted state. A single tau molecule begins roughly at band 0 and ends roughly at the next band 0. The overlap between successive molecules is probably small, so that the maximum repeat (68 nm) is close to the maximum molecular length (all springs extended). This is consistent with the observation that the separation of the bands changes nearly proportionally to the overall repeat (Figs 1d, 3).

subunits and the intervening clefts which result from specific interactions between the molecules⁷. There may be several discrete classes of interactions, each with its own characteristic banding pattern. Their differences persist even when scaling to the same repeat length. By contrast, the patterns of tau are roughly superimposable after scaling in spite of a more than threefold range in spacings. This can only be explained if the molecules themselves show elastic behaviour. This is consistent with the flexibility of the paracrystals, and with the anticorrelation between repeat and thickness (Fig. 2b).

A working model of tau protein is shown in Fig. 4. It consists of several domains connected by springs. In the paracrystals different molecules are joined head-to-tail, with the same polarities, and with little overlap. The maximum length would be given by the maximum repeat, about 68 nm, in agreement with images of glycerol-sprayed rotary shadowed tau molecules (data not shown). If the wispy 'protofibrils' seen in some images (Fig. 1e) correspond to tau molecules, their width would be ~3–6 nm, roughly consistent with the axial ratio of tau deduced from the hydrodynamic behaviour⁹.

The most conspicuous property of tau is its flexibility and elasticity. This can be deduced independently of the details of the molecular structure. It is probably related to the high content of Pro, Gly, and hydrophilic residues⁴, similar to other proteins with elastic properties such as elastin¹⁰. This protein has been described as a 'liquid drop elastomer', and it was pointed out that this property would allow very thin protein filaments containing globular domains to show elastic behaviour¹¹. Such an analysis seems appropriate for tau as well. It would argue against a role of tau as a stiff rod joining microtubules to each other or to other structures of the cell with an invariant spacing, but in favour of the tau providing a pliable surface coat on microtubules. Tau could confer flexibility to microtubules and/or provide elements of the cytoplasmic ground substance with the elasticity required to accommodate rearrangements and changes of direction or motion. This would be important, for example, in axoplasmic transport, where vesicles carried by translocator proteins must be able to move along microtubules¹², and any resulting perturbation must be removed after their passage. Proteins with elastic properties such as tau would also help to explain the saltatory particle movements observed in many cells, or the deformation of the marginal band microtubules of avian erythrocytes.

We thank P. Bendall and S. Fuller (EMBL) for their film

scanning facilities and P. Derr for photography. The project was supported by the Bundesministerium für Forschung und Technologie and the Deutsche Forschungsgemeinschaft. This study is part of thesis work to be submitted to Hamburg University (B.L., T.H.)

Received 2 May; accepted 13 June 1988.

- Weingarten, M. D., Lockwood, A. H., Hwo, S. Y. & Kirschner, M. W. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1858–1862 (1975).
- Kosik, K., Joachim, C. & Selkoe, D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4044–4048 (1986).

- Wood, J., Mirra, S., Pollock, N. & Binder, L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4040–4043 (1986).
- Lee, G., Cowan, N. & Kirschner, M. *Science* **239**, 285–288 (1988).
- Fellous, A., Francon, J., Lennon, A. M. & Nunez, J. *Eur. J. Biochem.* **78**, 167–174 (1977).
- Lindwall, G. & Cole, R. D. *J. biol. Chem.* **259**, 12241–12245 (1984).
- Doyle, B., Hukins, D., Hulmes, D., Miller, A. & Woodhead-Galloway, J. *J. molec. Biol.* **91**, 79–99 (1975).
- Caspar, D. L. D., Cohen, C. & Longley, W. *J. molec. Biol.* **41**, 87–107 (1969).
- Cleveland, D. W., Hwo, S.-Y. & Kirschner, M. W. *J. molec. Biol.* **116**, 227–247 (1977).
- Raju, K. & Anwar, R. A. *J. biol. Chem.* **262**, 5755–5762 (1987).
- Weis-Fogh, T. & Andersen, S. O. *Nature* **227**, 718–721 (1970).
- Allen, R. D. *et al. J. Cell Biol.* **100**, 1736–1752 (1985).
- Mandelkow, E.-M., Herrmann, M. & Rühl, U. *J. molec. Biol.* **185**, 311–327 (1985).
- Van Bruggen, E., van Breemen, J., Keegstra, W., Boekema, E. & van Heel, M. *J. Microsc.* **141**, 11–20 (1986).

Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S rRNA

Danesh Moazed, James M. Robertson & Harry F. Noller

Thimann Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064, USA

The elongation factors EF-Tu and EF-G interact with ribosomes during protein synthesis^{1,2}: EF-Tu presents incoming aminoacyl transfer RNA to the programmed ribosome as an EF-Tu-GTP-tRNA ternary complex and EF-G promotes translocation of peptidyl-tRNA and its associated messenger RNA from the A to the P site after peptidyl transfer. Both events are accompanied by ribosome-dependent GTP hydrolysis. Here we use chemical probes to investigate the possible interaction of these factors with ribosomal RNA in *E. coli* ribosomes. We observe EF-G-dependent footprints *in vitro* and *in vivo* around position 1,067 in domain II of 23S rRNA, and in the loop around position 2,660 in domain VI. EF-Tu gives an overlapping footprint *in vitro* at positions 2,655 and 2,661, but shows no effect at position 1,067. The 1,067 region is the site of interaction of the antibiotic thiostrepton³, which prevents formation of the EF-G-GTP-ribosome complex and is a site for interaction with the GTPase-related protein L11 (ref. 3). The universally conserved loop in the 2,660 region⁴ is the site of attack by the RNA-directed cytotoxins α -sarcin⁵ and ricin⁶, whose effects abolish translation and include the loss of elongation factor-dependent functions⁷ in eukaryotic ribosomes.

Complexes between EF-G and 70S ribosomes were stabilized either by GTP and fusidic acid⁸ or using the non-hydrolysable GTP analogues GMPPCP or GMPPNP (refs 9 and 10; Fig. 1

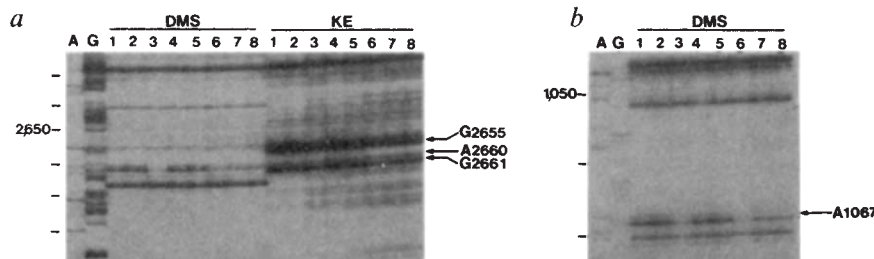
legend). These samples and appropriate controls were treated with dimethyl sulphate (DMS) and kethoxal as probes for A at N1, C at N3, G at N7 (DMS) and G at N1 and N2 (kethoxal). Modification of rRNA was monitored by primer extension (ref. 11 and Fig. 1). EF-G-dependent protection of G2655, A2660 and G2661 occurs when EF-G binding is stabilized either by fusidic acid (Fig. 1a, lanes 3 and 6) or by GTP analogues (Fig. 1a lanes 7 and 8). A1067 and A1069 are similarly protected (Fig. 1b). These effects are independent of the presence of cognate tRNA (compare lanes 3 and 6) or mRNA (not shown) and can be reproduced *in vivo* by DMS treatment of *Escherichia coli* cells that have been arrested during logarithmic growth by addition of fusidic acid (Fig. 2). Our results provide further support for the physiological relevance of the *in vitro* data.

Binding of the EF-Tu-GTP-aminoacyl-tRNA ternary complex to ribosomes can be stabilized by the antibiotic kirromycin¹². Bases in the 2,660 region are protected in a manner dependent on both EF-Tu and kirromycin (Fig. 3, lane 3): kirromycin alone (data not shown), or EF-Tu in the absence of kirromycin, either with GTP or GTP γ S (Fig. 3, lanes 2, 4), is insufficient. The chief EF-Tu-dependent protections are at G2655 and G2661, with weaker effects at A2660 and A2665 (Fig. 3, lane 3; Table 1). This contrasts with the results for EF-G, which show that the most strongly protected base is A2660 (Table 1). Thus both elongation factors interact either directly or indirectly with the α -sarcin loop in a similar, but not identical, way. The other main difference is the absence of any detectable EF-Tu-dependent effects in the 1,067 region. We fail to detect any protection using ternary complexes containing non-hydrolysable GTP analogues (for example, Fig. 3, lane 4; D.M. and J.M.R., unpublished results). This is despite the fact that such complexes promote binding of [¹⁴C]Phe-tRNA^{Phe} to 60–80% of ribosomes, as measured by standard filter-binding assays. In this case, protections due to bound tRNA are also observed (data not

Fig. 1 Autoradiographs showing EF-G-dependent protection of bases in the a, 2,660 and b, 1,067 regions of 23S rRNA against attack by dimethylsulphate and kethoxal (KE). A and G are dideoxy sequencing lanes and refer to the nucleotide sequence of 23S rRNA (ref. 4). Lane 1, 70S ribosomes + poly(U); lane 2, 70S + poly(U) + fusidic acid; lane 3, 70S + poly(U) + EF-G + GTP + fusidic acid; lane 4, 70S + poly(U) + tRNA^{Phe}; lane 5, 70S + poly(U) + tRNA^{Phe} + EF-G + GTP; lane 6, 70S + poly(U) + tRNA^{Phe} + EF-G + GTP + fusidic acid; lane 7, 70S + poly(U) + tRNA^{Phe} + EF-G + guanylyl-5'-imidodiphosphate (GMPPNP); lane 8, 70S + poly(U) + tRNA^{Phe} + EF-G + guanylyl-methylene-diphosphate (GMPPCP).

Positions of chemical attack are determined by primer extension using a set of synthetic DNA primers spaced at 200–250 nucleotide intervals along the entire 16S and 23S rRNA chains (ref. 11, and D.M. and H.F.N., unpublished results).

Methods *E. coli* MRE600 70S ribosomes were prepared as described²³, but with three washing steps using 0.5M NH₄Cl. EF-G was generously supplied by J. Bodley or prepared as described in ref. 24. Complexes of ribosomes and EF-G were formed by incubating 100 pmol EF-G with 40 pmol 70S ribosomes and 5 μ g poly(U) in 100 μ l 80 mM potassium cacodylate (pH 7.2), 10 mM MgCl₂, 50–100 mM NH₄Cl, 1 mM dithiothreitol, 0.5 mM EDTA, 0.5 mM GTP, 1 mM fusidic acid for 20 min at 20 °C and then for 5 min on ice. Other concentrations were as follows: tRNA^{Phe}, 0.8 μ M; GMPPNP, 0.5 mM; GMPPCP, 0.5 mM. Ribosomes were incubated with poly(U) for 5 min at 37 °C and with tRNA^{Phe} for another 5 min at 37 °C before adding EF-G. Chemical modification was by addition of DMS (2 μ l, 1:6 dilution in ethanol) or kethoxal (5 μ l, 37 mg ml⁻¹ in water), followed by incubation at 37 °C for 10 min. Reactions were stopped and the RNA extracted and resuspended as described¹¹. Primer extension and gel electrophoresis were performed as before^{11,25}.



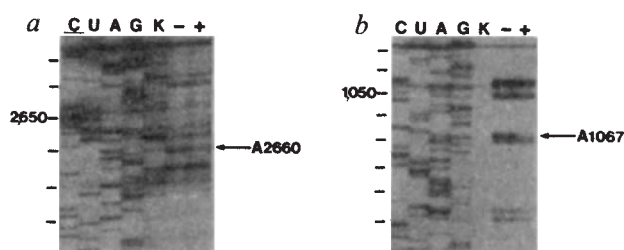


Fig. 2 *In vivo* protection of bases against DMS modification in 23S rRNA from *E. coli* cells arrested during logarithmic growth by treatment with fusidic acid. K, RNA from unmodified cells; -, RNA from untreated cells; +, RNA from cells treated with fusidic acid. *a* and *b* show the 2,660 and 1,067 regions respectively. Other abbreviations are as described in Fig. 1 legend.

Methods. *E. coli* MRE600 cells were grown to log phase in 150 ml LB (1% bactotryptone, 0.5% yeast extract, 1% NaCl) by shaking at 37 °C. Fusidic acid was added to one of the samples (to 1 mM) and incubation continued until growth stopped (~5 min). The cells were then poured over 100 g crushed ice and spun for 20 min at 8,000 r.p.m. in a GS3 rotor at 4 °C. Each cell pellet was resuspended in 5 ml ice-cold LB in 40 ml polycarbonate tubes. Chemical modification was performed by addition of 100 µl DMS (1:6 dilution in ethanol) followed by shaking for 5–7 min at 37 °C. The reaction was stopped by addition of 35 ml ice-cold 25 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 0.5 M β-mercaptoethanol. The cells were pelleted by spinning for 10 min at 8,000 r.p.m. in an SS34 rotor at 4 °C. The cell pellets were washed with 40 ml ice-cold 25 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 100 mM KCl, resuspended in 10 ml of the same buffer and opened by sonication. The cell lysate was spun twice for 10 min at 15,000 r.p.m. in an SS34 rotor at 4 °C. The second supernatant was spun for 3.5 h at 45,000 r.p.m. in a Ti50 rotor. The ribosome pellet was resuspended in 300 µl 0.3 M sodium acetate, 2 mM EDTA, 0.5% SDS and extracted three times with phenol and twice with chloroform. The RNA was then precipitated and resuspended as described¹¹.

shown), suggesting that the EF-Tu-ribosome interaction in complexes using GTP analogues is unstable under our experimental conditions. Finally, the EF-Tu-dependent effects seen here are unlikely to be due to tRNA, because tRNA fails to give any of these effects, whether it is bound to the P site or to the A site either enzymatically or non-enzymatically (Fig. 3, lanes 1 and 2, and unpublished experiments).

EF-G-dependent protection of A1067 and A1069 adds to increasing evidence that this region of 23S rRNA is involved in EF-G-related interactions. Thiostrepton, an antibiotic that specifically inhibits EF-G-ribosome interactions¹⁴, binds to an isolated fragment of 23S rRNA containing the 1,067 region; methylation of the 2'-hydroxyl of A1067 provides resistance to thiostrepton in strains of *Streptomyces azureus*³; EF-G itself has

Table 1 Bases in 23S ribosomal RNA protected from chemical probes by elongation factors

Nucleotide position	Relative band intensities		
	EF-Tu (kirromycin)	EF-G (fusidic acid)	EF-G (GMPPNP or GMPPCP)
A1067	0.95 (±0.08)	0.37 (±0.05)	0.47 (±0.03)
A1069	1.01 (±0.05)	0.74 (±0.08)	0.75 (±0.04)
G2655	0.38 (±0.04)	0.34 (±0.13)	0.48 (±0.06)
A2660	0.65 (±0.08)	0.28 (±0.08)	0.36 (±0.02)
G2661	0.48 (±0.03)	0.38 (±0.11)	0.50 (±0.05)
A2665	0.65 (±0.07)	0.96 (±0.02)	0.98 (±0.05)

Band intensities were measured by densitometry of autoradiographs and normalized to intensities of control bands. The values presented here are the averages of 3 or 4 experiments (columns 1 and 2) or 2 experiments (column 3), showing calculated standard deviations. Full reactivity is defined as 1.00.

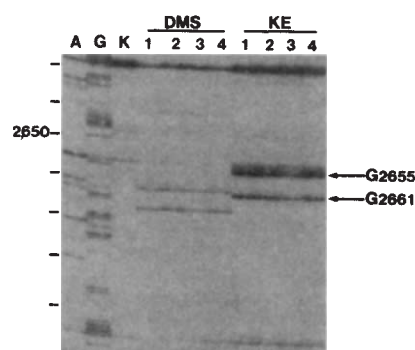


Fig. 3 Protection of bases in the 2,660 region of 23S rRNA caused by binding of EF-Tu to 70S ribosomes. Lane 1, 70S + poly(U) + tRNA^{Phe}; lane 2, 70S + poly(U) + tRNA^{Phe} + [¹⁴C]Phe-tRNA^{Phe} + EF-Tu + GTP; lane 3, 70S + poly(U) + tRNA^{Phe} + [¹⁴C]Phe-tRNA^{Phe} + EF-Tu + GTP + kirromycin; lane 4, 70S + poly(U) + tRNA^{Phe} + [¹⁴C]Phe-tRNA^{Phe} + EF-Tu + GTPγS. Abbreviations are described in Fig. 1 legend.

Methods. EF-Tu-GDP (gift from F. Jurnak; ref. 29) was converted to EF-Tu-GTP by incubation of 50 pmol EF-Tu-GDP in 20 µl 80 mM potassium cacodylate, pH 7.2, 5 mM MgCl₂, 100 mM NH₄Cl, 5 mM dithiothreitol, 0.5 M GTP (or GTPγS), 1.5 µM EF-Ts for 30 min at 20 °C. *E. coli* tRNA^{Phe} (Subriden, Rollingbay, WA) was charged to 1,200 pmol per absorbance unit at 260 nm with [¹⁴C]phenylalanine (513 mCi mmol⁻¹, Amersham) as described³⁰. 25 pmol [¹⁴C]Phe-tRNA^{Phe} in 5 µl 80 mM potassium cacodylate, pH 7.2, 5 mM MgCl₂, 100 mM NH₄Cl was incubated with 20 µl EF-Tu complex (above). This was added to the same buffer (75 µl) containing 20 pmol 70S ribosomes, 5 µg poly(U), 40 pmol tRNA^{Phe}. Ribosomes were incubated with poly(U) and tRNA^{Phe} for 20 min at 37 °C before addition of ternary complex. Concentration of kirromycin (lane 3) was 100 µM. Binding of [¹⁴C]Phe-tRNA^{Phe} to ribosomes was tested by millipore filtration²⁶. 80–91% of ribosomes bound [¹⁴C]Phe-tRNA^{Phe} in an EF-Tu-dependent manner. Chemical modification was performed for 5 min at 37 °C as described in Fig. 1 legend. As in the experiments of Fig. 1 and 2, the entire 16S and 23S rRNA sequences were scanned by primer extension to identify sites of modification.

been cross-linked to 23S rRNA in the vicinity of A1067 (ref. 15) and the ribosomal proteins L11 and L7/L12, which have been implicated in translational GTPase-related functions^{16,17} bind in³ or near²² the 1,067 region of 23S rRNA respectively. Our results support the idea that thiostrepton acts by interfering with the interaction of EF-G with 23S rRNA.

Both EF-G and EF-Tu protect bases in the universally conserved loop around position 2,660 of 23S rRNA. This loop is the site of action of both α-sarcin, which cleaves the G2661–A2662 phosphodiester linkage⁵, and ricin, which removes adenine 2,660 in a N-glycosidase reaction⁶. Both of these cytotoxins abolish protein synthesis in eukaryotic ribosomes⁷. Our results provide evidence that this is due to alteration of the structure of a region of rRNA that interacts with EF-1 and EF-2, the eukaryotic counterparts of EF-Tu and EF-G²⁷. Protection of bases in the α-sarcin loop by both EF-Tu and EF-G is in agreement with earlier competitive binding studies^{18–21} and with immuno-electron microscopy mapping (reviewed in ref. 28), which implied that these two factors bind to the ribosome at overlapping sites. As both factors are involved in ribosome-dependent GTPase functions, these results suggest a specific function for this universally conserved loop of ribosomal RNA. Finally, our results indicate the existence of an EF-G-mediated functional interaction between domains II and VI, two regions that are widely separated in the primary and secondary structures of 23S rRNA. These findings support the suggestion that RNA is involved in virtually every aspect of the translational process⁴, an idea that is consistent with the remarkable conservation of primary and secondary structure in the corresponding regions of the RNA.

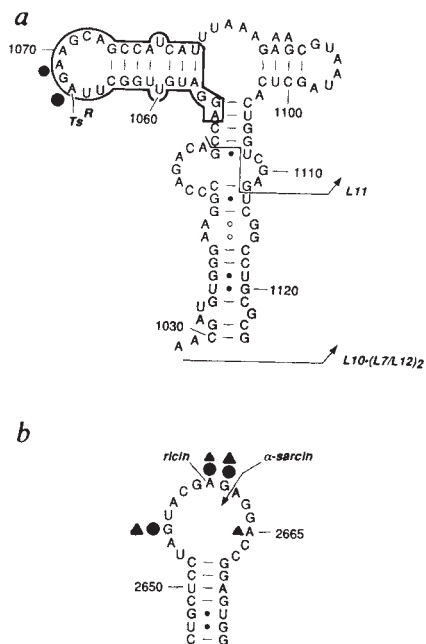


Fig. 4 Regions of *E. coli* 23S rRNA (a, b) showing positions where binding of elongation factors causes protection of bases against chemical probes. Symbols (●) and (▲) indicate protection by EF-G and EF-Tu, respectively. The site of cleavage by α -sarcin⁵, depurination by ricin⁶, and the binding sites for ribosomal protein L11 (1,052–1,112) (ref. 3) and ribosomal protein complex L10-(L7/L12)₂ (1,028–1,124) (ref. 22) are also shown. The boxed RNA fragment has been cross-linked to EF-G by diepoxybutane¹⁵. Methylation of the 2'-hydroxyl of A1067 provides resistance to thiostrepton³.

We are indebted to Drs James Bodley and Frances Jurnak for generously providing the purified elongation factors used in these studies. We thank Dr Heinz Wolf for a gift of kirromycin, and Bryn Weiser for Fig. 4. This work was supported by NIH (H.F.N.).

Received 26 May; accepted 15 June 1988

1. Kaziro, Y. *Biochim. biophys. Acta* **505**, 95–127 (1978).
2. Lucas-Lenard, J. & Lipmann, F. A. *Rev. Biochem.* **40**, 409–448 (1971).
3. Schmidt, F. J., Thompson, J., Lee, K., Dijk, J. & Cundliffe, E. *J. biol. Chem.* **266**, 12301–12305 (1981).
4. Noller, H. F. A. *Rev. Biochem.* **53**, 119–162 (1984).
5. Endo, Y. & Wool, I. G. *J. biol. Chem.* **257**, 9054–9060 (1982).
6. Endo, Y., Mitsui, K., Motizuki, M. & Tsurugi, K. *J. biol. Chem.* **262**, 5908–5912 (1987).
7. Fernandez-Puentes, C. & Vazquez, D. *FEBS Lett.* **78**, 143–146 (1977).
8. Bodley, J. W., Zieve, F. J. & Lin, L. *J. biol. Chem.* **245**, 5662–5667 (1970).
9. Hershey, J. W. B. & Monro, R. E. *J. molec. Biol.* **18**, 68–76 (1966).
10. Eckstein, F., Kettler, M. & Parmeggiani, A. *Biochem. biophys. Res. Commun.* **45**, 1151–1158 (1971).
11. Moazed, D., Stern, S. & Noller, H. F. *J. molec. Biol.* **187**, 399–416 (1986).
12. Wolf, H., Chinali, G. & Parmeggiani, A. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4910–4914 (1974).
13. Yokosawa, H., Inoue-Yokosawa, N., Arai, K., Kawakita, M. & Kaziro, Y. *J. biol. Chem.* **248**, 375–377 (1973).
14. Bodley, J. W., Lin, L. & Highland, J. H. *Biochim. biophys. Acta* **91**, 1406–1411 (1970).
15. Sköld, S. *Nucleic Acids Res.* **11**, 4923–4932 (1983).
16. Maassen, J. A. & Möller, W. *J. biol. Chem.* **253**, 2777–2783 (1978).
17. Möller, W. in *Ribosomes* (eds Nomura, M., Tissières, A. & Lengyel, P.) 711–731 (Cold Spring Harbor Laboratory, New York, 1974).
18. Richman, N. & Bodley, J. W. *Proc. natn. Acad. Sci. U.S.A.* **69**, 688–689 (1972).
19. Cabrer, B., Vazquez, D. & Modolell, J. *Proc. natn. Acad. Sci. U.S.A.* **69**, 733–736 (1972).
20. Miller, D. *Proc. natn. Acad. Sci. U.S.A.* **69**, 752–755 (1972).
21. Richter, D. *Biochim. biophys. Acta* **46**, 1850–1856 (1972).
22. Beauclerk, A. A., Cundliffe, E. & Dijk, J. *J. biol. Chem.* **259**, 6559–6563 (1984).
23. Moazed, D., Van Stolk, J., Douthwaite, S. & Noller, H. F. *J. molec. Biol.* **191**, 483–493 (1986).
24. Robertson, J. M., Urbanke, C., Chinali, G., Wintermeyer, W. & Parmeggiani, A. *J. molec. Biol.* **189**, 653–662 (1986).
25. Stern, S., Moazed, D. & Noller, H. F. *Meth. Enzym.* (in the press).
26. Nirenberg, M. & Leder, P. *Science* **145**, 1399–1407 (1964).
27. Moldave, K. A. *Rev. Biochem.* **54**, 1109–1150 (1985).
28. Lake, J. A. *Rev. Biochem.* **54**, 507–530 (1985).
29. Louie, A., Ribeiro, N. S., Reid, B. R. & Jurnak, F. *J. biol. Chem.* **259**, 5010–5016 (1984).
30. Robertson, J. M. & Wintermeyer, W. *J. molec. Biol.* **151**, 57–79 (1981).

Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis

Dipankar Sen & Walter Gilbert

Department of Cellular and Developmental Biology, Harvard University, Cambridge, Massachusetts 02138, USA

We have discovered that single-stranded DNA containing short guanine-rich motifs will self-associate at physiological salt concentrations to make four-stranded structures in which the strands run in parallel fashion. We believe these complexes are held together by guanines bonded to each other by Hoogsteen pairing. Such guanine-rich sequences occur in immunoglobulin switch regions¹, in gene promoters^{2,3}, and in chromosomal telomeres⁴. We speculate that this self-recognition of guanine-rich motifs of DNA serves to bring together, and to zipper up in register, the four homologous chromatids during meiosis.

We found this self-recognition of certain guanine-rich DNA sequences in the course of experiments examining the immunoglobulin switch regions. Using synthesized DNA fragments, we had begun mobility shift experiments seeking specific proteins that might bind to the switch regions. But the oligonucleotides aggregated and moved to novel positions on the gels; we investigated these complexes to discover that they were parallel sets of four strands.

The genes for the immunoglobulin heavy chains undergo switch recombinations to bring different constant regions next to the rearranged variable region exons during the differentiation of B lymphocytes to plasma cells¹. The switch-regions, the sites of these recombinations, lie 5' to the constant-region genes and are stretches of repetitive DNA 1–10 kilobases (kb) in size. They consist largely of tandem repeats of GC-rich DNA which are 20–50 base pairs (bp) long and have various degrees of similarity. The coding-strand consensus sequences⁵ contain typically 65–70% purines, with guanines making up 50% of the total. Several repetitive motifs occur, such as GGGGT and GAGCT, and the conserved sequence (G)GGGGAGCTGGGG, found in Sγ1, Sγ2b and Sγ3.

We synthesized oligomers, 20–49 bases long, corresponding to the IgG switch regions. The guanine-rich strands are shown in the legend to Fig. 1. When these oligomers were stored in buffer containing monovalent salt, each spontaneously generated a complex of markedly lower electrophoretic mobility. Figure 1, lane 1, shows this complex (M4) for one of the oligomers, M. The pyrimidine-rich complementary strands, in contrast, show no tendency to form analogous complexes.

These complexes formed in a concentration-dependent way, and this fact, together with their low electrophoretic mobility, suggested that they were inter-strand aggregates rather than internally folded, base-paired single strands. Kinased mixtures of the oligomer and complex were readily separated from one another by preparative gel-electrophoresis (Fig. 1, lane 3).

To determine the structure of such a complex, we first examined its ability to bind its complementary strand. Figure 1 shows the results of mixing labelled samples of the complex of the oligomer M and its complementary strand R in various ratios and analysing portions on a non-denaturing 6% polyacrylamide gel. Four new bands, of even slower mobility than the complex, appear in lanes 4–8 (a portion of the complex breaks down to form the Watson-Crick dimer '2'). As increasing amounts of the complementary strand are added, the label in all four upper bands coalesces into the fourth and uppermost band, which we thus infer to be a M4R4 product. Lanes 10–14 show a mirror experiment with the label in the complementary strand. The

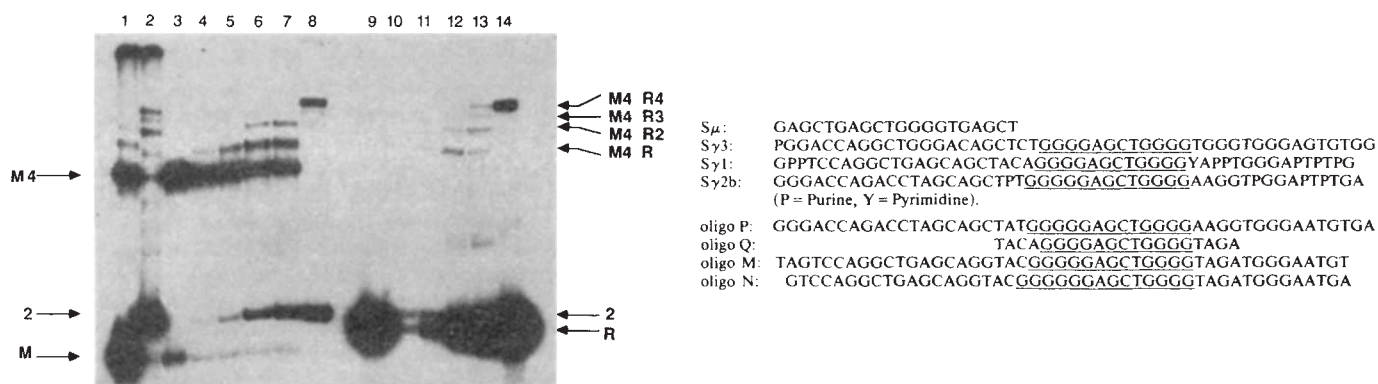


Fig. 1 Left, non-denaturing polyacrylamide-gel mobilities of the oligomer M and its complexes. M: oligomer M monostrand; M4: tetra-strand of M; R: complementary strand to M; 2: Watson-Crick double strand of M and R; M4R-M4R4: M4 complexed with one to four strands of R. Consensus sequences⁵ of switch-region repeat units are shown on the right, where P is the 49b repeat of $S_{\gamma 2b}$; Q contains the conserved motif GGGGAGCTGGGG; M and N are derived from $S_{\gamma 1}$, except that N contains an extra guanine residue, to give a six-guanine stretch. **Methods.** Oligomers were made using a Milligen Autogen 6500 DNA synthesizer. Complexes were formed by storing the oligonucleotides at 100–500 $\mu\text{g ml}^{-1}$ in 10 mM Tris pH 8, 90 mM NaCl, 1 mM EDTA buffer (TNE) for 24–48 h at 4 °C and then isolated on preparative gels. Association reactions were carried out in 10 μl of TNE buffer at 20 °C, for 30 min. Lane 1, 1 ng total of oligomer M and its complex M4; lane 2, 1 ng labelled M and M4 added to 5 ng cold complementary strand R; lane 3, 0.4 ng gel-purified M4; lanes 4–8, 0.4 ng labelled M4, mixed with 0.1, 0.3, 1.0, 3.0 and 10.0 ng respectively of unlabelled R; lane 9, 3 ng labelled complementary strand R; lanes 10–14, 0.1 ng unlabelled M4, mixed with 0.1, 0.3, 1.0, 3.0 and 10.0 ng respectively of labelled R. Samples were electrophoresed through a 6% polyacrylamide gel in 50 mM TBE buffer⁶ at 20 °C.

formation of the four upper bands, each containing the complementary strand R, shows that the complex of M is a tetramer (termed M4), the upper bands thus representing the binding of 1–4 complementary strands to M4.

To elucidate the structure of the complex, the guanines of gel-purified M and M4 were methylated with dimethyl sulphate⁶. Piperidine-treated samples, run on a sequencing gel, revealed a complete methylation-protection of the sequence GGGGAGCTGGGG (Fig. 2a, lanes M: 1 and 2). The unavailability of the N-7s of these guanines for methylation suggests their involvement in Hoogsteen bonding⁷. Figure 2a also shows a similar methylation-protection pattern for all of the complexes M4, N4, P4 and Q4. In P4 the protected region spans 21 bases with a total of 15 guanines protected, extending over the sequence GGGGAGCTGGGGAAGGTGGG. But the GGGs at the 3' end of M and N are not protected; the sequence of the other bases must also be relevant to the ability of the complex to extend.

The protection patterns of M4, N4 and P4 clearly demonstrate that the orientation of the four strands is *parallel*. No antiparallel orientation could completely protect the N-7s of every guanine within the motifs.

X-ray fibre diffraction studies and model-building on polyG (refs 8, 9) have suggested that the ribo-homopolymer could form a four-stranded structure, with the guanines Hoogsteen bonded, as in Fig. 2b. Our results demonstrate that short guanine-rich motifs, in a natural DNA sequence, can form stable four-stranded structures and are able to maintain their four-strandedness even when complexed to complementary strands. Figure 3 shows the structures we predict: M4R4 represents the uppermost electrophoretic band of Fig. 1, with the complex M4 bound to four complementary strands R.

Might this self-recognition and four-strand formation in parallel orientation by guanine-rich strands of DNA have a function *in vivo*?

We propose that these four-stranded regions, which we will call G4-DNA, have a role in the pairing of homologous chromosomes during meiosis. In the early stages of meiosis, after DNA synthesis, four copies of the DNA double helix, two pairs of sister chromatids, are brought together. In the stages of prophase 1 partially condensed homologous chromosomes appear precisely aligned, side by side, as a prelude to eventual exchanges of genetic information between the homologues (reviewed in refs 10, 11).

Synapsis is sometimes observed to begin at telomeres (reviewed in refs 4, 12) to generate 'bouquet' structures, although in other instances, it is initiated at intra-chromosomal loci^{13,14}. The association of telomeres is consistent with our hypothesis, because all non-mitochondrial telomeres sequenced to date contain repeated guanine-rich motifs that are added to the chromosome subsequent to replication⁴. Furthermore, the single-strands containing the guanine-rich motifs are usually longer than their complementary strands and thus overhang them by 12–16 bases (reviewed in ref. 15).

Henderson *et al.*¹⁶ have recently shown that guanine-rich synthetic oligomers from various telomeres can form guanine-mediated fold-back structures at low temperatures. But judging from the similarity of the telomere sequences to the switch sequences, we predict that telomere sequences will also form G4-DNA at physiological salt concentrations. Although we would not rule out a role for proteins in telomere-telomere interactions *in vivo*, we propose that the primary driving force for their association may come from this intrinsic property of parallel aggregation by self-recognizing guanine-rich sequences. Oka and Thomas¹⁷ examined the aggregation of macronuclear DNA from *Oxytricha* and showed that it was dependent upon a telomeric overhang of --TTTGGGGTTTGGGG being single stranded. We interpret this observation as involving the formation of G4-DNA structures: four Watson strands interacting in contrast to Oka and Thomas' interpretation of two Watson and two Crick strands being involved.

Figure 4 sketches our proposal that G4-DNA elements tie together the four chromatids at meiosis. Once joined at the telomeres, or elsewhere, the tight association of the homologues (zippering up) proceeds via a recognition of 'self' and subsequent four-strand formation in parallel fashion of specific sequences along the chromatid. A number of loci along the chromosome, each containing a guanine-rich motif of unique size and sequence, together constitutes a chromosomal fingerprint to facilitate recognition and association. The guanine-rich strands of the motifs might be made accessible to each other by local unwinding or by the process described by Mirkin *et al.*¹⁸ in which homopurine-homopyridine tracts of a duplex DNA undergo a superhelix-induced structural transition to a novel DNA conformation, the H form, leaving a polypurine strand free from base pairing. Most simply, however, the cell could control G4-DNA formation by synthesizing both melting proteins that would bind to C-rich DNA as well as G4-binding

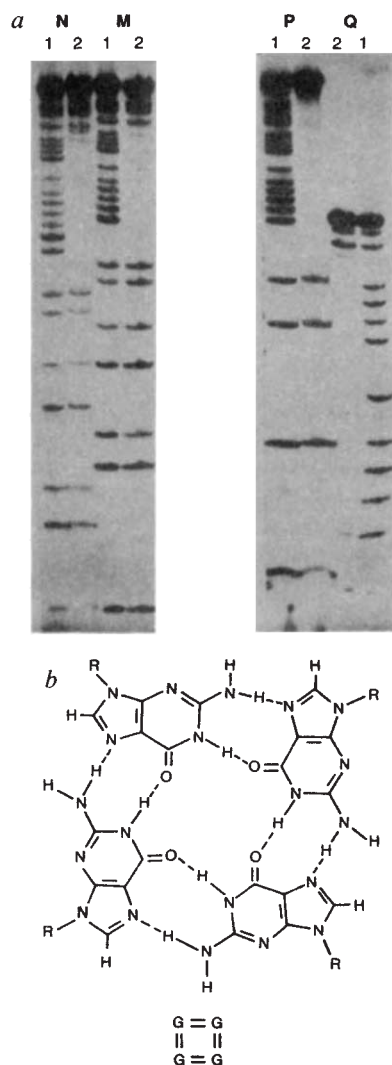


Fig. 2 *a*, N-7 methylation patterns of oligomers M, N, P and Q (lanes labelled 1), and their respective complexes (lanes labelled 2). Methylation reactions were adapted from Maxam and Gilbert⁶. Kinased DNA was treated with dimethyl sulphate and electrophoresed on a 6% non-denaturing acrylamide gel. Bands of complex and monomer were detected by autoradiography, eluted into TNE, precipitated, treated with piperidine, and analysed on a 20% sequencing gel. *b*, Model structure of a Hoogsteen hydrogen-bonded guanine tetrad^{8,9}.

proteins that would stabilize G4-DNA, irrespective of sequence. Removing both types of proteins, or even synthesizing G4-melting proteins, would disrupt the pairing.

G4-DNA forming regions may replicate late as the last step before pairing. Hotta and Stern¹⁹ observed that there is a late synthesis, during zygotene, of ~0.3% of the DNA, GC-rich, that is required in order that meiosis can proceed.

The pairing of non-homologues too might be initiated by G4-DNA formation from near-identical guanine motifs, but

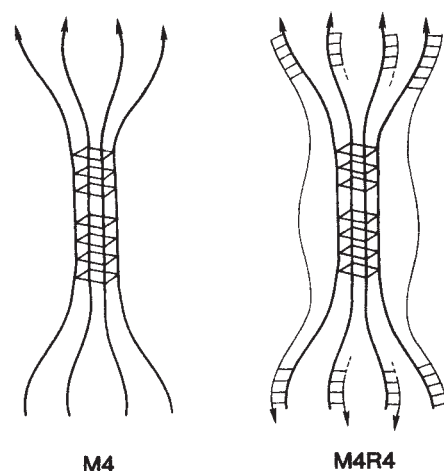


Fig. 3 Schematic model for the complexes M4 and M4R4.

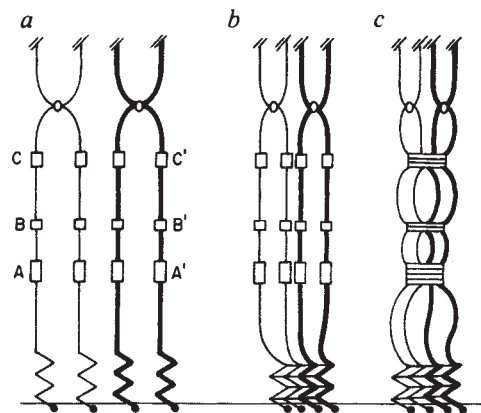


Fig. 4 A scheme for the participation of guanine-rich telomeres and internal chromosomal motifs in homologue recognition during meiotic prophase. Telomeres are shown by zigzagged lines, and guanine-rich 'recognition' motifs as A, B and C (with counterparts A', B' and C'). *a*, Homologous chromosomes are shown with their telomeres attached to the nuclear envelope. *b*, The four telomeres associate, by G4-DNA formation, to generate 'bouquets'. *c*, Recognition and synapsis occur mediated by stretches of G4-DNA along the homologues.

would not be successful over the entire length of dissimilar chromosomes. There is some cytological evidence for such loose pairing²⁰.

In summary, we attribute the interaction of four double helices at meiosis to an underlying four-fold recognition event mediated by short stretches of guanines, in a like-with-like interaction.

We thank Dr Matthew Meselson for discussions. This work was supported in part by the NIH. We thank Milligen for the DNA synthesizer.

Received 3 May; accepted 14 June 1988.

- Shimizu, A. & Honjo, T. *Cell* **36**, 801-803 (1984).
- Evans, T., Schon, E., Gora-Maslak, G., Patterson, J. & Efstratiadis, A. *Nucleic Acids Res.* **12**, 8043-8058 (1984).
- Kilpatrick, M. W., Torri, A., Kang, D. S., Engler, J. A. & Wells, R. D. *J. biol. Chem.* **261**, 11350-11354 (1986).
- Blackburn, E. H. & Szostak, J. W. *A. Rev. Biochem.* **53**, 163-194 (1984).
- Nikaido, T., Yamawaki-Kataoka, Y. & Honjo, T. *J. biol. Chem.* **257**, 7322-7329 (1982).
- Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1977).
- Hoogsteen, K. *Acta Crystallogr.* **12**, 822-823 (1959).
- Arnott, S., Chandrasekaran, R. & Marttila, C. M. *Biochem. J.* **141**, 537-543 (1974).
- Zimmerman, S. B., Cohen, G. H. & Davis, D. R. *J. molec. Biol.* **92**, 181-192 (1975).

- John, B. *Chromosoma* **54**, 295-325 (1976).
- Maguire, M. P. *J. theor. Biol.* **106**, 605-615 (1984).
- Ashley, T. & Wagenaar, E. B. *Can. J. Genet. Cytol.* **16**, 61-76 (1974).
- Gillies, C. B. *Trav. Lab. Carlsberg* **40**, 135 (1975).
- Holm, P. B. *Trav. Lab. Carlsberg* **42**, 103 (1977).
- Weiner, A. M. *Cell* **52**, 155-157 (1988).
- Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I. & Blackburn, E. H. *Cell* **51**, 899-908 (1987).
- Oka, Y. & Thomas, C. A. *Nucleic Acids Res.* **15**, 8877-8898 (1987).
- Mirkin, S. M. *et al. Nature* **330**, 495-497 (1987).
- Hotta, Y. & Stern, H. *J. molec. Biol.* **55**, 337-355 (1971).
- Moens, P. B. *Cold Spring Harb. quant. Symp.* **49**, 699-705 (1984).